Absenteeism at work

Abstract

Absenteeism at the workplace is a pervasive issue that can significantly impact productivity, team dynamics, and overall organizational performance. In this study, we delve into a comprehensive dataset collected from a corporate environment to explore the underlying patterns and identify potential predictors of absenteeism. Our objective is to gain insights that can inform targeted interventions and policies aimed at reducing absenteeism and improving workplace well-being.

Currently, absenteeism at work is considered as the absence of the worker from his/her job, either as a result of delays or anticipated service, and one of the main problems faced by companies and associations. Absenteeism at work is something undesirable for companies and associations, negatively impacting productivity and the organizational climate itself, and generating financial losses for companies.

Absenteeism is a habitual pattern of absence from a duty or obligation without good reason. Generally, absenteeism is unplanned absences. If a workplace exhibits a high degree of absenteeism there is a problem. It has been viewed as an indicator of poor individual performance, as well as a breach of an implicit contract between employee and employer.

The absence of an employee can generate a gap on a certain day, in a production line, in the elaboration of a project or decision making, etc...

Work relationships are becoming more and more trust-oriented, and conservative contracts are being replaced with more agile ones, in which the employee is responsible for accounting for the working time. This liberty may lead to unregulated absenteeism and may reflect poorly on an employee's candidature, even if the absent hours can be accounted for with genuine reasons. This can significantly undermine healthy working relationships. Moreover, unregulated absenteeism can also harm work productivity.

The database was created with records of absenteeism at work from July 2007 to July 2010 at a courier company in Brazil. The dataset allows for various attribute notations and attribute deletions, or a modification of the attribute type (categorical, integer, or real) dependent on the purpose of the search.

Let's analyze the data and observe some interesting characteristics of employee profiles, we often imagine that the unruliest employee is the one who is most lacking, or the employee with the most children, such data in this company demonstrate that often not everything follows a rule.

One of the key findings of our analysis is the identification of potential early indicators of absenteeism. By leveraging historical data, we develop predictive models that demonstrate promising accuracy in forecasting future absenteeism patterns. These models consider a combination of individual characteristics and workplace variables, providing a holistic view of the factors contributing to absenteeism. The implications of these findings extend beyond mere prediction, offering an opportunity for organizations to implement targeted preventive measures and support systems.

At the end, we will develop an algorithm based on the profile of employees who have no absences, if there are new hires, Human Resources can use the algorithm to help the hiring, whether you hire or not, if you have the profile of being a missing employee or not.

Introduction

Main problem of this project

Problem Statement: XYZ is a courier company. As we appreciate that human capital plays an important role in collection, transportation and delivery. The company is passing through the genuine issue of Absenteeism. The company has shared it dataset and requested to have an answer on the following areas:

1. What changes company should bring to reduce the number of absenteeism?

2. How much losses every month can we project in 2011 if same trend of absenteeism continues

The problem is that the work environment of today is more:

- Competitive
- Managers set unachievable business goals.
- have an elevated risk of becoming unemployed.

This can lead to an increase in pressure and stress on the employee. Those factors influence employee health, which is of course undesirable.

Techniques used

I am going to divide the whole project in to 8 parts:

1.) Define and categorize problem statement.

2.) Visualize the data 3.) Data cleaning

4.) Perform Exploratory Data Analysis

5.) Model Building

- Feature Reduction
- Naïve Bayes implementation
- Bayesian Belief Network
- Decision Tree
- KNN
- Neural Network

6.) Evaluate and compare Model performances and choose the best model.

7.) Hypertune the selected model

8.) Produce sample output with tuned model.

The main contribution of this project lies in its thorough analysis of absenteeism patterns and predictors, the development of predictive models, and the provision of actionable insights for organizations to proactively manage absenteeism and enhance overall workplace satisfaction.

The organization of the rest of the project can follow a structured framework, encompassing various sections to provide a comprehensive understanding of the absenteeism study.

Related Work

Methods	Results	Reference	Year
Linear Discriminant Analysis	Applied LDA to identify latent topics	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
Principal Component Analysis	PCA for dimensionality reduction in absenteeism analysis	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
SVD	Achieved 99% accuracy in predicting absenteeism	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
Decision Tree	Mean Absolute Error: 0.04 Mean Squared Error: 0.02	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
KNN	Mean Absolute Error: 0.04 Root mean Squared Error: 0.16	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
Neural Network	Achieved 98% accuracy	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
K-fold cross validation	Average Accuracy: 0.99594	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
Bayesian Belief Network	Explored uncertainty in absenteeism predictions, providing probabilistic estimates	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
Naïve Bayes implementation	Accuracy score: 99.3 %	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003
overfitting or underfitting	it indicates overfitting	Faressayah. (2003). Predict employee absenteeism from work. Retrieved from https://www.kaggle.com/code/faressayah/predict-employee-absenteeism-from-work	2003

Methodology

Methods I used on my dataset:

1. Linear Discriminant Analysis (LDA):
- LDA is a dimensionality reduction that focuses on maximizing the distance between the means of different classes while minimizing the spread within each class. LDA is particularly useful for problems involving multiple classes, and it provides a linear decision boundary for classification tasks.

2. Principal Component Analysis (PCA):
- PCA is a dimensionality reduction technique that transforms data into a new coordinate system, capturing the most significant variance in the data. It is often used to simplify complex datasets and identify key patterns by projecting data onto a lower-dimensional space.

3. Singular Value Decomposition (SVD):
- SVD is a linear algebra technique used for matrix factorization. It decomposes a matrix into three other matrices, representing the singular values, left singular vectors, and right singular vectors.

4. Naïve Bayes Implementation:
- Naïve Bayes assumes that features are conditionally independent given the class label. Naïve Bayes is effective for text classification and other applications.

5. Bayesian Belief Network:
- A Bayesian Belief Network (BBN) is a graphical model that represents probabilistic relationships among a set of variables. It uses Bayesian inference to update probabilities based on new evidence.

6. Decision Tree:
- A Decision Tree is a supervised machine learning algorithm that recursively partitions data based on the most significant features. It is used for classification and regression tasks, providing a tree-like structure where each node represents a decision based on a feature.

7. K-Nearest Neighbors (KNN):
- KNN is used for classification and regression. It classifies a data point based on the majority class among its k nearest neighbors in the feature space.

8. Neural Network:

- A Neural Network consists of interconnected nodes organized into layers. Neural networks can learn complex relationships in data and are widely used for tasks such as image recognition and natural language processing.

9. K-fold Cross Validation:

- K-fold Cross Validation is a model evaluation technique that partitions the dataset into k subsets. It iteratively trains and tests the model k times, using a different subset as the test set in each iteration. It helps assess the model's performance and generalization to unseen data.

10. Receiver Operating Characteristic (ROC):

- ROC is a graphical representation of a binary classification model's performance. It plots the true positive rate against the false positive rate at various threshold settings. The area under the ROC curve (AUC) is a common metric for evaluating the overall performance of a classification model.

Proposed Model

Project Phases and Methods:

❖ Preprocessing:
- Objective: Prepare the raw data for analysis by addressing missing values, handling outliers, and ensuring consistency.
- Methods:
- Data Cleaning: Impute missing values or remove incomplete records.
- Outlier Detection and Handling: Identify and address outliers using techniques such as Z-score or IQR.
- Data Transformation: Standardize or normalize numerical features to a common scale.
- Encoding Categorical Variables: Convert categorical variables into numerical format (e.g., one-hot encoding).

❖ Feature Selection:
  - Objective: Identify the most relevant features that contribute to the target variable.
  - Methods:
    - Correlation Analysis: Identify highly correlated features.

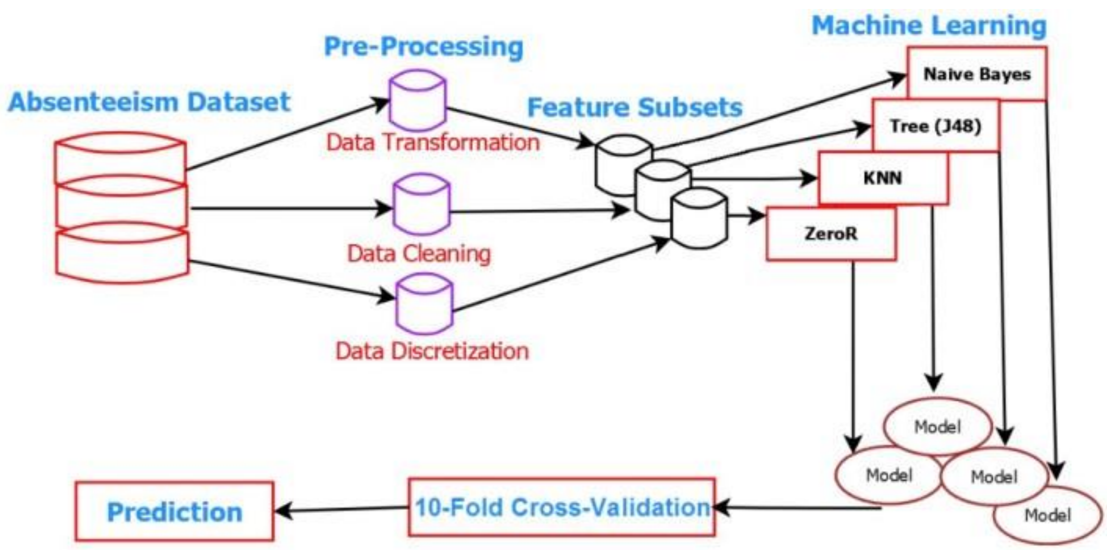
❖ Feature Reduction:
- Objective: Reduce the dimensionality of the dataset to enhance model efficiency and interpretability.
- Methods:
- Principal Component Analysis (PCA): Transform the data into a new set of uncorrelated variables (principal components).
- Linear Discriminant Analysis (LDA): Find the linear combinations that best separate different classes.

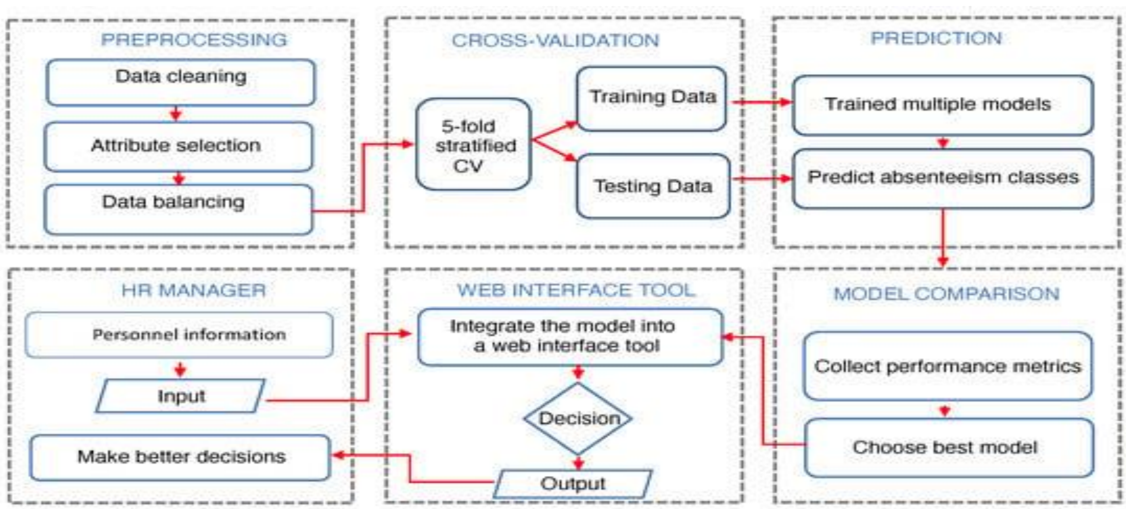
❖ Classification/Regression Methods:

- Objective: Build predictive models based on the preprocessed and feature-selected/reduced dataset.
- Methods:
- Logistic Regression: For binary classification problems.
- Decision Trees: Non-linear models that partition the feature space.
- Random Forest: Ensemble of decision trees for improved accuracy and robustness.
- Support Vector Machines (SVM): Effective for both classification and regression tasks.
- K-Nearest Neighbors (KNN): Classify data points based on the majority class among their neighbors.
- Neural Networks: Deep learning models for complex tasks.
- Linear Regression: For predicting continuous target variables.

❖ Evaluation Metrics:

- Objective: Assess the performance of the models and choose the best-suited model.
- Common Classification Metrics:
  - Accuracy, Precision, Recall, F1 Score.
  - Confusion Matrix.
  - ROC-AUC Curve.
- Common Regression Metrics:
  - Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared.
- Cross-Validation: Use techniques like K-fold cross-validation to estimate model performance on unseen data.





Results and discussion

Data Set Description

The Absenteeism-data.csv is a dataset based on an existing study about absenteeism in a workplace. It contains information regarding the employee ID, Reason for Absence, Transportation Expense, Distance to Work, Age, Daily Workload Average, Body Mass Index, Education, Children, Pets, and Absenteeism Time (Hours).

The relevant libraries are imported, the raw data is loaded using the pandas library, and a copy is made to keep the original dataset unharmed.

The human resources department of a company has analyzed the data of its employees and discovered that there is a problem with absenteeism.

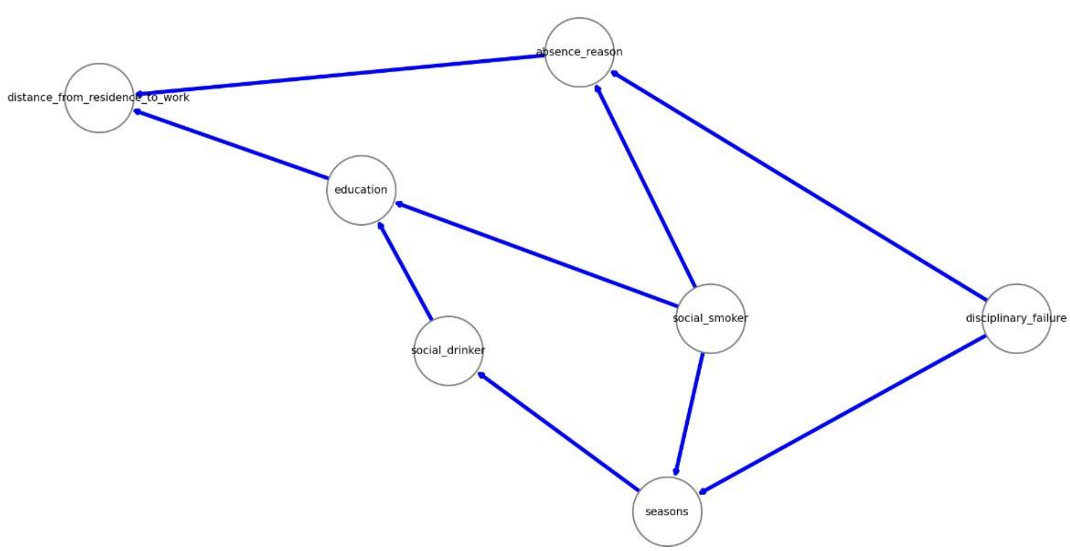
They have collected data and want to make a model that correctly classifies employees based on whether they miss more than four hours a month or not. This will help them to make future hires, and even keep track of their current employees.

Results of each method

1. **Naïve Bayes implementation**

    - Accuracy score: 99.32432432432432 %

2. **Bayesian Belief Network**



3. **Decision Tree**

- Mean Absolute Error: 0.03648648648648649
- Mean Squared Error: 0.02459459459459946
- Root Mean Squared Error: 0.15682663866382712

4. **Neural Network**

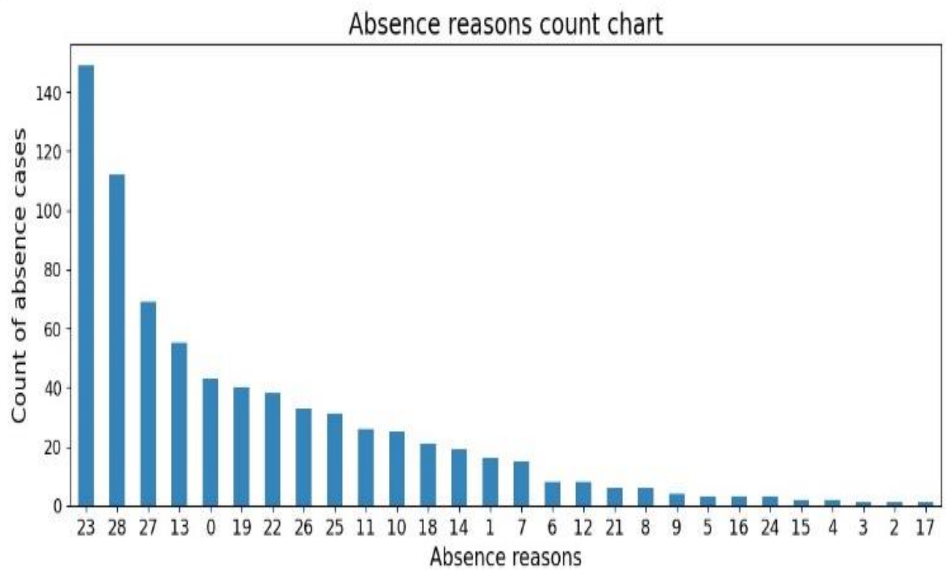
```
# Training the neural network on the training data.
model.fit(X_train, y_train, epochs=10, batch_size=32, validation_data=(X_test, y_test))
```

```
Epoch 1/10
19/19 [==============================] - 3s 43ms/step - loss: 1.0076 - accuracy: 0.8936 - val_loss: 0.3701 - val_accuracy: 0.9730
Epoch 2/10
19/19 [==============================] - 0s 14ms/step - loss: 0.5936 - accuracy: 0.9003 - val_loss: 0.2701 - val_accuracy: 0.9459
Epoch 3/10
19/19 [==============================] - 0s 15ms/step - loss: 0.3474 - accuracy: 0.9307 - val_loss: 0.1686 - val_accuracy: 0.9730
Epoch 4/10
19/19 [==============================] - 0s 16ms/step - loss: 0.1678 - accuracy: 0.9459 - val_loss: 0.1475 - val_accuracy: 0.9392
Epoch 5/10
19/19 [==============================] - 0s 18ms/step - loss: 0.1098 - accuracy: 0.9611 - val_loss: 0.1453 - val_accuracy: 0.9527
Epoch 6/10
19/19 [==============================] - 0s 19ms/step - loss: 0.0996 - accuracy: 0.9595 - val_loss: 0.0537 - val_accuracy: 0.9797
Epoch 7/10
19/19 [==============================] - 0s 18ms/step - loss: 0.0836 - accuracy: 0.9679 - val_loss: 0.0324 - val_accuracy: 0.9865
Epoch 8/10
19/19 [==============================] - 0s 18ms/step - loss: 0.0895 - accuracy: 0.9662 - val_loss: 0.0517 - val_accuracy: 0.9797
Epoch 9/10
19/19 [==============================] - 0s 17ms/step - loss: 0.0991 - accuracy: 0.9628 - val_loss: 0.0200 - val_accuracy: 0.9932
Epoch 10/10
19/19 [==============================] - 0s 16ms/step - loss: 0.0435 - accuracy: 0.9831 - val_loss: 0.0272 - val_accuracy: 0.9797
<keras.src.callbacks.History at 0x153b968bb10>
```

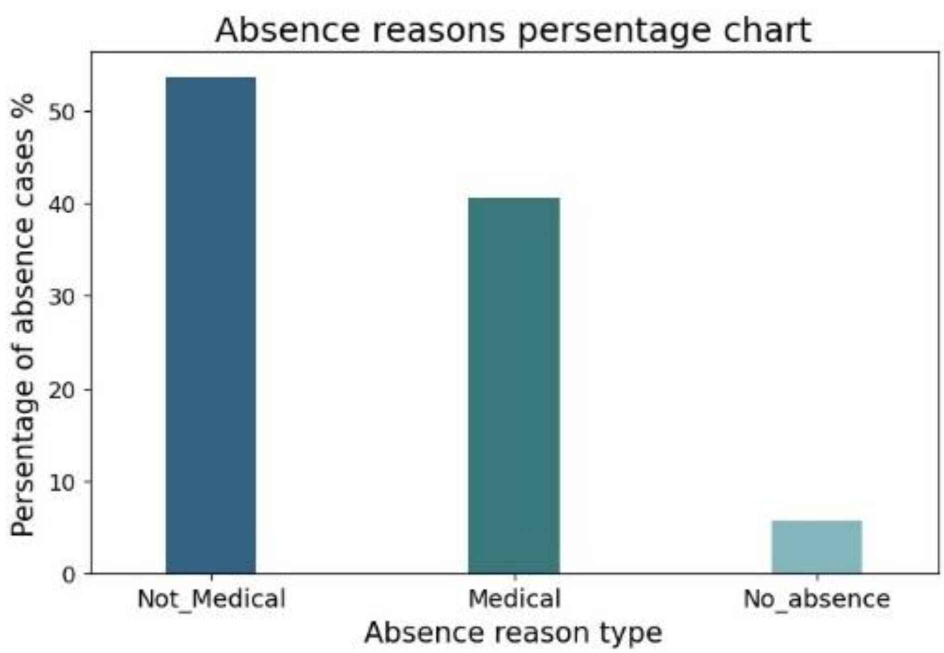
```
# Evaluate the performance of the model on the test set.
test_loss, test_accuracy = model.evaluate(X_test, y_test)
print(f'Test Accuracy: {test_accuracy}')
```

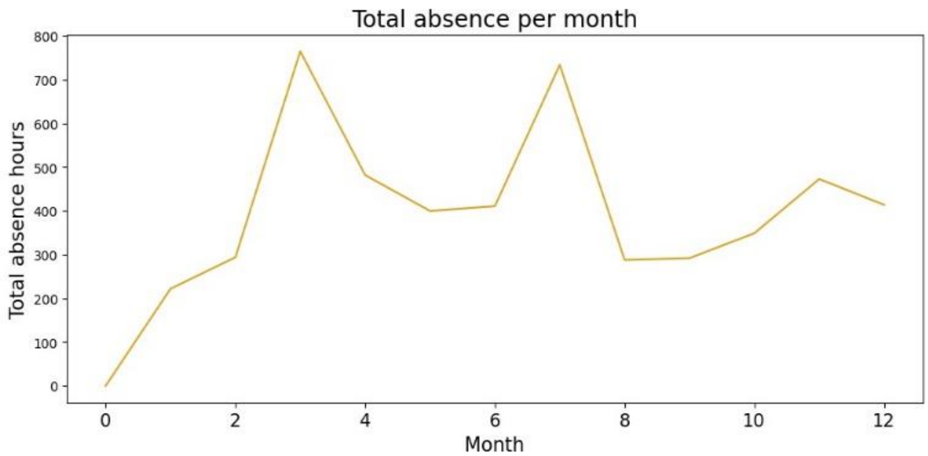
```
5/5 [==============================] - 0s 8ms/step - loss: 0.0272 - accuracy: 0.9797
Test Accuracy: 0.9797297120094299
```

Results of the preprocessing phase

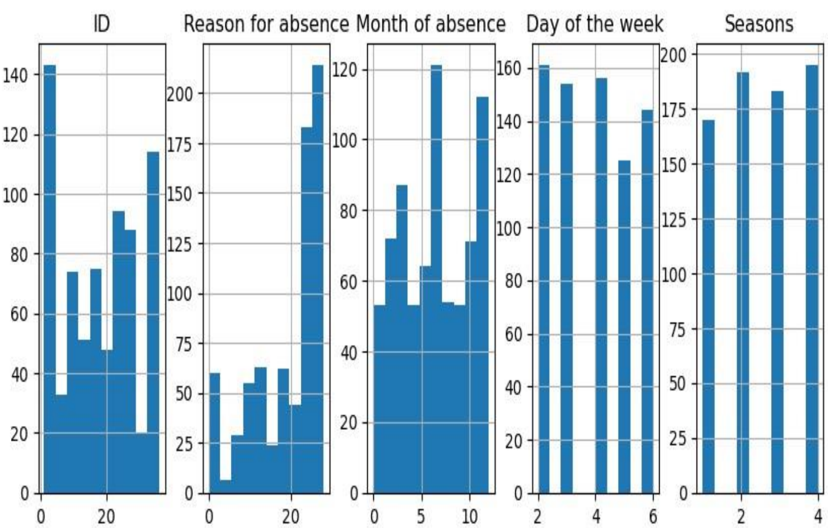


From above I can conclude that the most frequent reason for absences is blood donation which is represented by number (23) with 149 absence case.

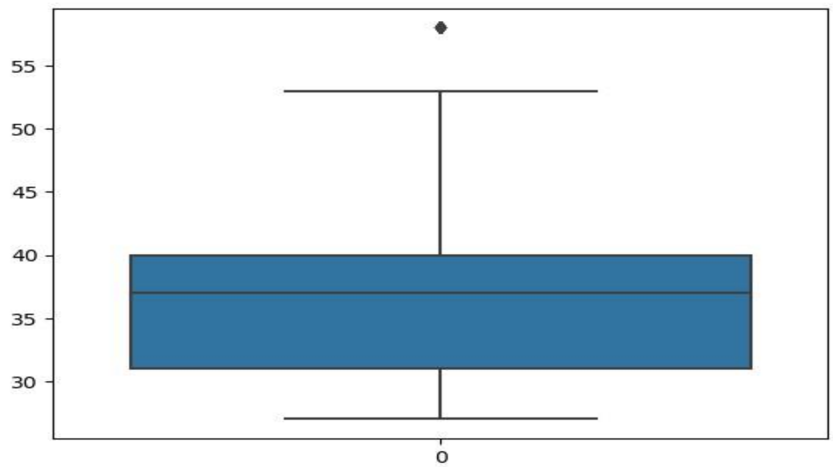


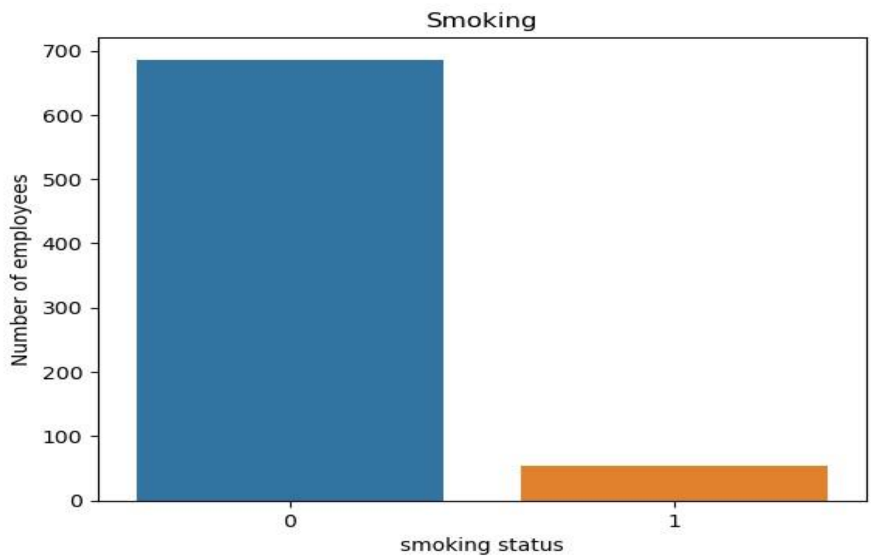


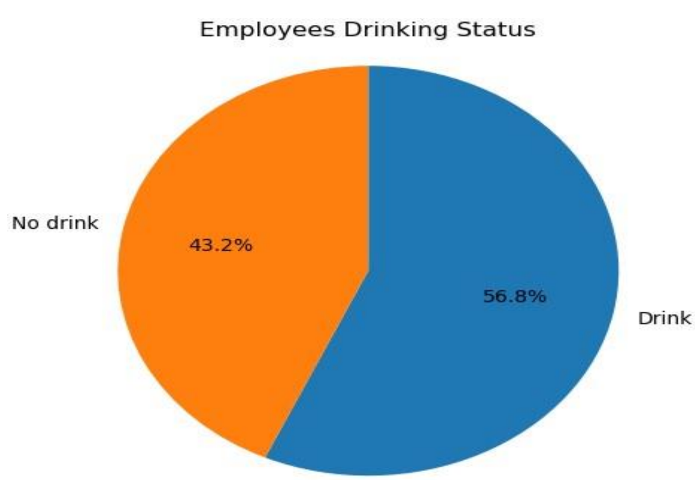
from above i can see that: March has the higher absence rate with total 765 hours. January has the lowest absence rate with total 222 hours.



```
# create a boxplot of the age
sns.boxplot(df.Age)
plt.show()
```







❖ Min, Max, Mean, Standard Deviation

	ID	Reason for absence	Month of absence	Day of the week	Seasons	Transportation expense	Distance from Residence to Work	Service time	Age
count	740.000000	740.000000	740.000000	740.000000	740.000000	740.000000	740.000000	740.000000	740.000000
mean	18.017568	19.216216	6.324324	3.914865	2.544595	221.329730	29.631081	12.554054	36.450000
std	11.021247	8.433406	3.436287	1.421675	1.111831	66.952223	14.836788	4.384873	6.478772
min	1.000000	0.000000	0.000000	2.000000	1.000000	118.000000	5.000000	1.000000	27.000000
25%	9.000000	13.000000	3.000000	3.000000	2.000000	179.000000	16.000000	9.000000	31.000000
50%	18.000000	23.000000	6.000000	4.000000	3.000000	225.000000	26.000000	13.000000	37.000000
75%	28.000000	26.000000	9.000000	5.000000	4.000000	260.000000	50.000000	16.000000	40.000000
max	36.000000	28.000000	12.000000	6.000000	4.000000	388.000000	52.000000	29.000000	58.000000

❖ Variance

```
df.var()
```

```
ID                                  121.467891
Reason for absence                   71.122335
Month of absence                     11.808068
Day of the week                       2.021159
Seasons                               1.236168
Transportation expense             4482.600197
Distance from Residence to Work     220.130291
Service time                         19.227115
Age                                  41.974493
Work load Average/day              1525.536440
Hit target                           14.283208
Disciplinary failure                  0.051201
Education                             0.453249
Son                                   1.206678
Social drinker                        0.245767
Social_smoker                         0.067739
Pet                                   1.737805
Weight                              165.977113
Height                               36.421159
Body mass index                      18.365101
Absenteeism time in hours           177.715510
```

❖ Skewness

```
skewness = df.skew()
print("Skewness:", skewness)
kurtosis_value = df.kurtosis()
print("Kurtosis:", kurtosis_value)
```

```
Skewness: ID                                 0.016606
Reason for absence                -0.915312
Month of absence                   0.069369
Day of the week                    0.102440
Seasons                           -0.038532
Transportation expense             0.396189
Distance from Residence to Work    0.312083
Service time                      -0.004720
Age                                0.697703
Work load Average/day              0.961457
Hit target                        -1.261708
Disciplinary failure               3.952270
Education                          2.108953
Son                                1.086465
Social drinker                    -0.273327
Social_smoker                      3.290333
Pet                                2.735715
Weight                             0.017001
Height                             2.566060
Body mass index                    0.305046
Absenteeism time in hours          5.720728
```

❖ Kurtosis

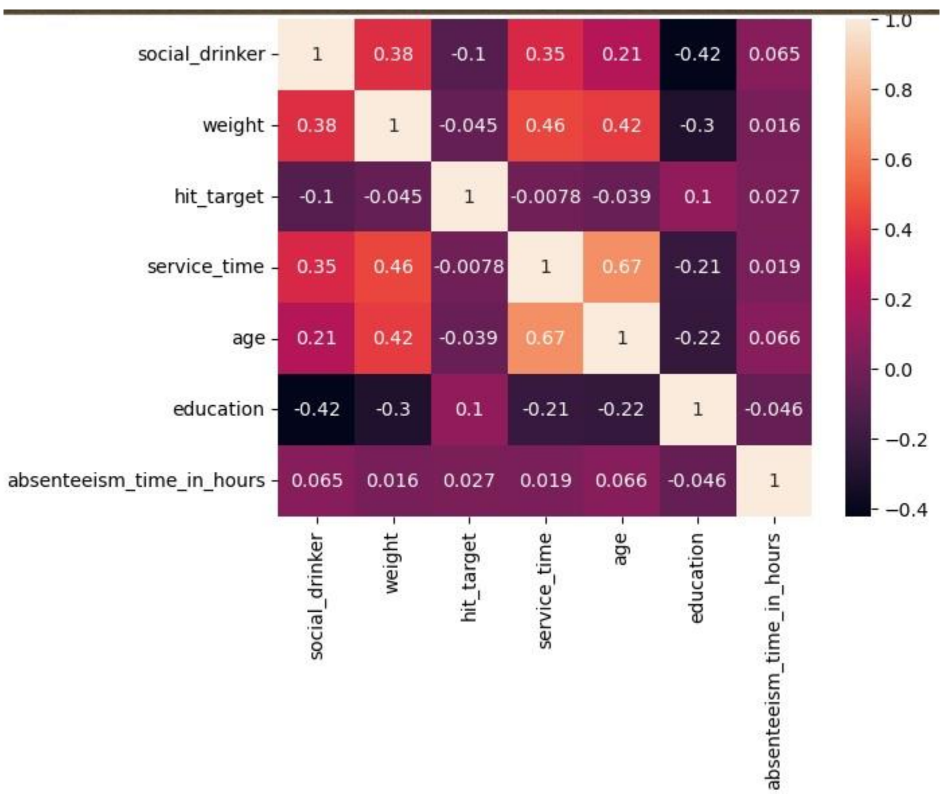
```
Kurtosis: ID                                -1.251818
Reason for absence                -0.259925
Month of absence                  -1.254967
Day of the week                   -1.286406
Seasons                           -1.345190
Transportation expense            -0.318291
Distance from Residence to Work   -1.261683
Service time                       0.683111
Age                                0.431613
Work load Average/day              0.618188
Hit target                         2.419042
Disciplinary failure              13.657345
Education                          2.988465
Son                                0.748326
Social drinker                    -1.930517
Social_smoker                      8.850204
Pet                                9.674827
Weight                            -0.913928
Height                             7.317235
Body mass index                   -0.314375
Absenteeism time in hours         38.777307
```

❖ Correlation

```
df.corr()
```

	ID	Reason for absence	Month of absence	Day of the week	Seasons	Transportation expense	Distance from Residence to Work
ID	1.000000	-0.064245	-0.000043	0.034468	0.098495	-0.224163	-0.486160
Reason for absence	-0.064245	1.000000	-0.083858	0.116319	-0.117925	-0.119381	0.161831
Month of absence	-0.000043	-0.083858	1.000000	-0.006528	0.407770	0.137525	-0.003887
Day of the week	0.034468	0.116319	-0.006528	1.000000	0.046493	0.033988	0.118026
Seasons	0.098495	-0.117925	0.407770	0.046493	1.000000	0.036995	-0.063108
Transportation expense	-0.224163	-0.119381	0.137525	0.033988	0.036995	1.000000	0.262183
Distance from Residence to Work	-0.486160	0.161831	-0.003887	0.118026	-0.063108	0.262183	1.000000
Service time	-0.272704	0.048425	-0.062862	0.021252	-0.010904	-0.349887	0.131730
Age	0.040899	-0.078608	-0.001520	0.004459	-0.012089	-0.227542	-0.145886
Work load Average/day	0.092457	-0.123472	-0.169989	0.015646	0.150439	0.005438	-0.068677
Hit target	0.018789	0.088943	-0.460453	0.030986	-0.061154	-0.080193	-0.013865
Disciplinary failure	0.004502	-0.545054	0.107946	-0.015120	0.151766	0.109222	-0.056527

❖ Heat map



❖ **Z-test**

- Z-Score : 2.5562240203851085
- Critical Z-Score : 1.6448536269514722
- Reject Null Hypothesis
- p-value : 0.005290748561472158

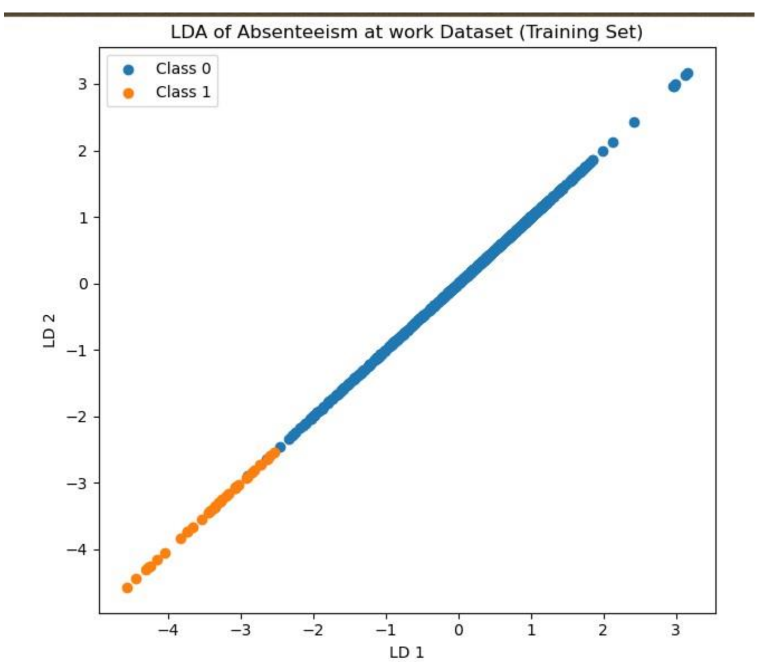
❖ **ANOVA**

- ANOVA Statistic: 2283.4528652366844
- P-value: 0.0
- Reject the null hypothesis. There is a significant difference between the means.

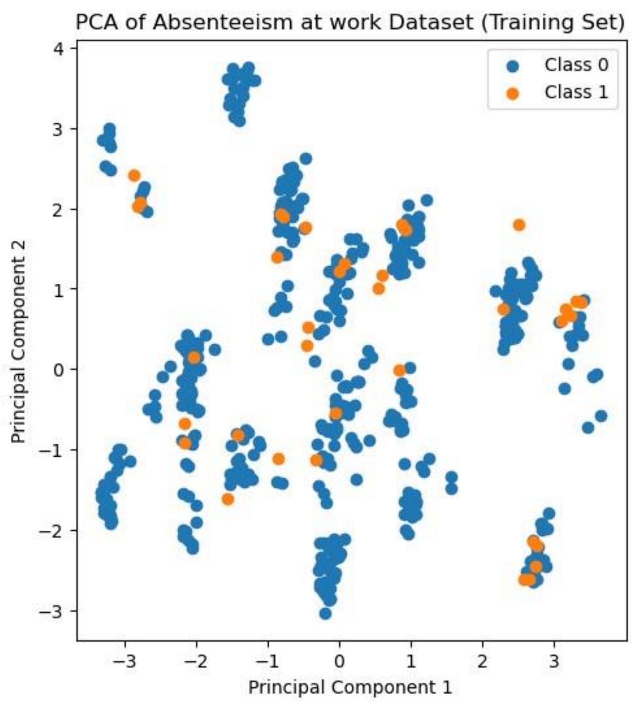
❖ **Chi- square Test**
- Pearson Chi-square (1.0) = 1.9885
- p-value = 0.1585
- Cramer's phi = 0.05

Feature Reduction results

Linear Discriminate Analysis (LDA)



Principle Component Analysis (PCA)



Singular Value Decomposition (SVD)

- Accuracy: 0.9932432432432432

- Write the phase of classification / Regression methods results with Tables and Figures.

split Dataset to (80% training and 20% testing)

Data splitting

```
In [24]: # Function for training and evaluating a classifier
def train_and_evaluate(classifier, X_train, X_test, y_train, y_test):
    classifier.fit(X_train, y_train)
    y_pred = classifier.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    return accuracy
```

```
In [35]: X = df_clean.drop("disciplinary_failure",axis=1)
y= df_clean["disciplinary_failure"] #We will predict disciplinary failure
```

```
In [36]: # We will choose almost 20% of dataset as test size.
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2 ,random_state=1)
print('Number of rows in the total set: {}'.format(df_clean.shape[0]))
print('Number of rows in the training set: {}'.format(X_train.shape[0]))
print('Number of rows in the test set: {}'.format(X_test.shape[0]))
```

```
Number of rows in the total set: 740
Number of rows in the training set: 592
Number of rows in the test set: 148
```

-
-

K-fold cross validation and the average accuracy

K-fold cross validation

```
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import KFold
from sklearn.ensemble import RandomForestClassifier

# Assuming df is your DataFrame and X, y are features and labels
X = df_clean.drop('disciplinary_failure', axis=1)
y = df_clean['disciplinary_failure']

# Initialize your model
model = RandomForestClassifier()

# Set up k-fold cross-validation (e.g., with k=5 folds)
kfold = KFold(n_splits=5, shuffle=True, random_state=42)

# Perform cross-validation and get accuracy scores
cv_scores = cross_val_score(model, X, y, cv=kfold, scoring='accuracy')
```

```
# Display accuracy scores for each fold
print("Accuracy Scores for Each Fold:")
print(cv_scores)
# Calculate and display the average accuracy
average_accuracy = cv_scores.mean()
print("\nAverage Accuracy:", average_accuracy)
```

```
Accuracy Scores for Each Fold:
[1.         1.         0.98648649 1.         0.99324324]

Average Accuracy: 0.995945945945946
```

- `Average Accuracy: 0.995945945945946`

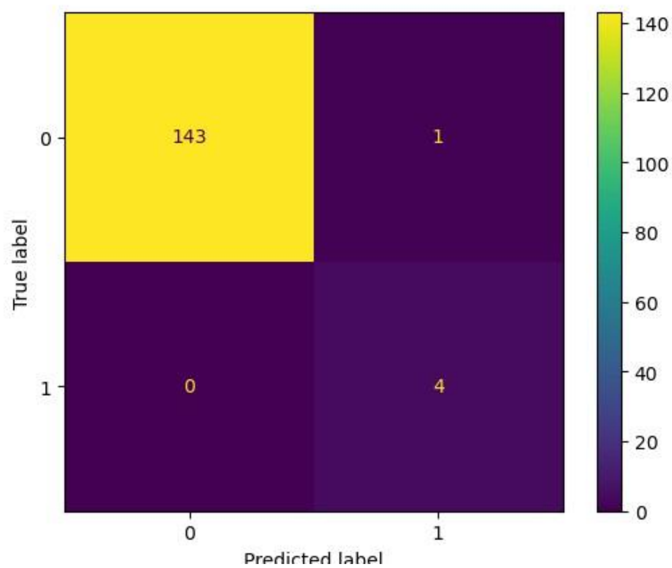
Confusion Matrix (Accuracy, Error rate, Precision, Recall, F-measure, and ROC) for each classifier

	Accuracy	Precision	Recall	Error Rate	F1 Score	ROC
LogisticRegression()	100%	100%	100%	0%	100%	99.65%
GaussianNB()	100%	95%	100%	0%	97%	99.7%
DecisionTreeRegressor()	100%	100%	100%	0%	100%	100%
LinearDiscriminantAnalysis()	99%	88%	100%	1%	94%	50%

```
[55]: print('confusion_matrix: \n', format(confusion_matrix(y_test, predictions)))

confusion_matrix:
 [[143   1]
 [  0   4]]
```

```
[38]: from sklearn.metrics import ConfusionMatrixDisplay
ConfusionMatrixDisplay.from_predictions(y_test, predictions)
plt.show()
```



The model performs better on the training set than the test set, it indicates overfitting.

Conclusion and future work

The business environment today is more competitive than it used to be. This leads to increased pressure in the workplace. And so, it is reasonable to expect an increase in stress levels of the employees. This can lead to a variety of problems ranging from physical to mental illnesses. But we looked at this problem from the point of view of the person responsible for improving productivity in the company.

I can see after working on this dataset that employees with low performance cause a vital lose for organizations and the absenteeism consider to be one of the factors that affect performance So, understanding the causes of absenteeism may power the organization with a competitive advantages tool and open the area of research for computer and human resources fields.

We explored whether a person presenting certain characteristics is expected to be away from work at some point in time. In other words, and after working on this dataset I was able to find answers to these questions:

1. How far do they live from the workplace?
2. Do they have children or pets?
3. If they do, how many?
4. Do they have a higher education degree?

In addition, I achieved the main Target from this project which is to build a Machine learning model which can predict employee's absence with high accuracy and ensure the precision and recall are perfect. The analysis helped me to discover the factors and causes of employees absence using computerized technologies.

The project is consisted of 3 stages:

1. Data cleaning in python
2. Model building
3. evaluation and prediction

Future work direction and how you can achieve better results using other methods.

Future Work and Improvements:

Feature Engineering:

- Discuss the potential for extracting additional information from existing features or exploring interactions between features.

Data Balancing and Time Series Analysis:

- Address the need to handle class imbalances if present and explore time series analysis techniques if the dataset has a temporal component.

Advanced Regression Models and Data Quality:

- Propose trying more advanced regression models and revisiting data preprocessing steps, such as handling outliers and missing values.

Domain Knowledge Integration and External Validation:

- Stress the value of incorporating domain knowledge more effectively and validating models on external datasets to assess generalizability.

Experiment with Neural Networks and Collaboration with Domain Experts:

- Suggest experimenting with neural networks depending on the data complexity and stress the importance of collaborating with domain experts.

Ethical Considerations:

- Mention the need to consider ethical aspects of model predictions, especially in workplace contexts, and ensure fairness and transparency.

References

[1] Martiniano, A., Ferreira, R. P., Sassi, R. J., & Affonso, C. (2012). Application of a neuro fuzzy network in prediction of absenteeism at work. In Information Systems and Technologies (CISTI), 7th Iberian Conference on (pp. 1-4). IEEE.

[2] F. Gomide, M. Figueiredo, W. Pedrycz, A neural Fuzzy network: Structure and learning, Fuzzy Logic and Its Applications, Information Sciences and Intelligent Systems, Bien, Z. and Min, K., Kluwer Academic Publishers, Netherlands, 1998, pp. 177-186.

[3] J. M. Ivancevich. Predicting Absenteeism from Prior Absence and Work Attitudes. The Academy of Management Journal. Vol. 28, no. 1, p. 219-228, mar. 1985.

[4] J. H. Morris, J. D. Sherman, R. A. Snyder. Prediction of Absenteeism from Attitudes, Prior Absenteeism, and Performance, Personnel Review, Vol. 18 no. 1, p.16 - 22, 1989.8