EPOCH

Supervised Learning Algorithm

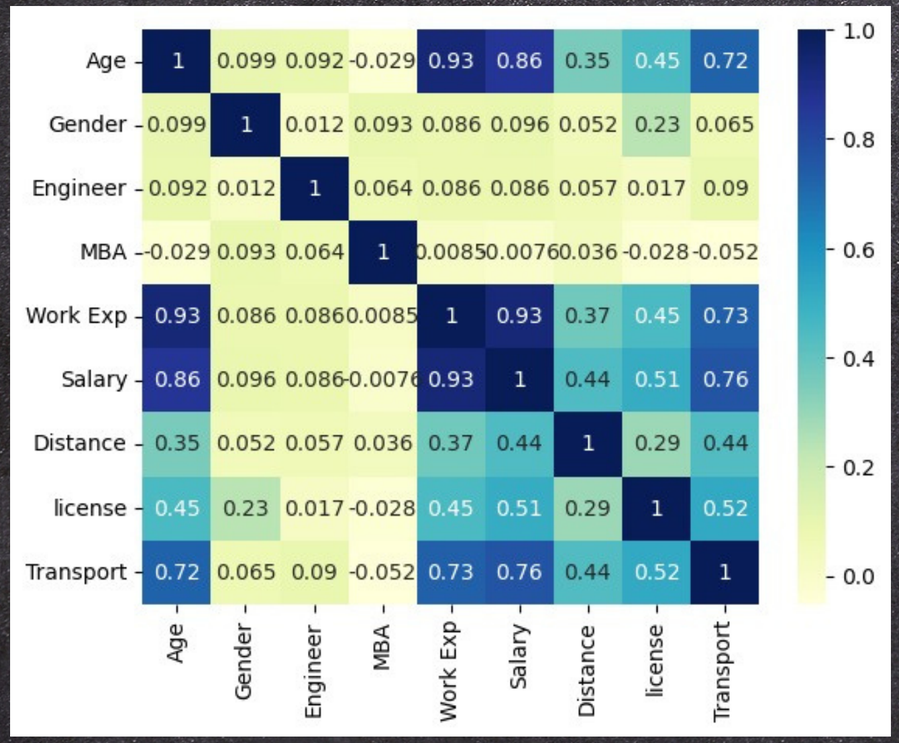
20/09/2023

The dataset combines employee transport preferences with personal and professional data, encompassing age, salary, and work experience. Our goal is to predict whether an employee will opt for a car as their mode of transport. Furthermore, we'll identify the most influential factors guiding this choice.

EXPLORATORY DATA ANALYSIS:

Rows	444
Columns	8
Discrete columns	1
Continuous columns	7
All missing columns	0
Missing observations	1
Complete Rows	443

CORELATION FACTOR



KEY CORRELATIONS

Variable A	Variable B	Correlation Coefficient
Age	Work Experience	Strong (0.93)
Age	Salary	Strong (0.86)
Salary	Work Experience	Strong (0.93)
Distance	Salary	Moderate (0.44)
License	Transport	Moderate (0.52)
Transport	Age	Strong (0.72)
MBA	Transport	Weak (-0.05)
Gender	License	Weak (0.23)
Distance	Transport	Moderate (0.44)

KEY INSIGHTS FROM CORRELATION ANALYSIS

Insight 1: Age, Work Experience, and Salary are Closely Related

- Strong positive correlation between these attributes.
- May indicate career progression.

Insight 2: Distance and Salary Influence Transportation Choices

- Moderate positive correlation.
- Salary may affect willingness to commute.

KEY :

```
from sklearn import preprocessing
label_encoder = preprocessing.LabelEncoder()
df['Transport']=df['Transport'].replace('Public Transport',0)
df['Transport']=df['Transport'].replace('2Wheeler',0)
df['Transport']=df['Transport'].replace('Car',1)
```

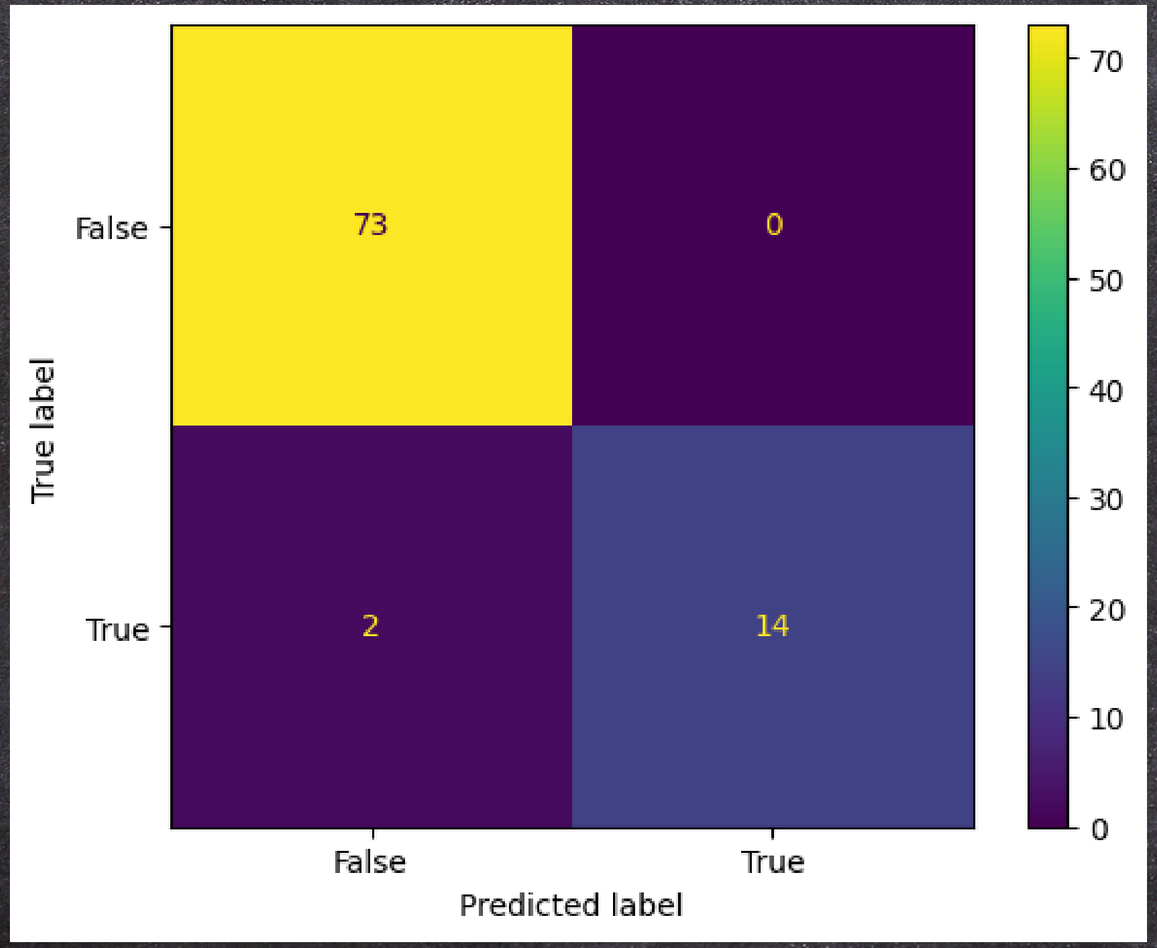
```
df=df.drop(columns='Engineer')
```

The code uses scikit-learn's LabelEncoder to convert 'Transport' categories into numeric values, where 'Public Transport' and '2Wheeler' are encoded as 0, and 'Car' is encoded as 1.

RANDOMFOREST

20/09/2023

CONFUSION MATRIX



- The model has a very low number of False Positives (0), which means it never misclassifies employees who do not prefer a specific transport mode as if they do.
- There are only (2)false negatives in the new matrix, indicating a low number of cases where the model incorrectly predicts the negative class when it should have predicted the positive class.
- This suggests a reasonably good ability to capture employees who actually prefer a specific transport mode.

PREDICTION ACCURACY

```
[66] from sklearn.ensemble import RandomForestClassifier
     classifier2= RandomForestClassifier(n_estimators= 10, criterion="entropy")
     classifier2.fit(x_train, y_train)
```

```
RandomForestClassifier(criterion='entropy', n_estimators=10)
```

```
[67] y_pred2= classifier2.predict(x_test)
```

```
accuracy = metrics.accuracy_score(y_test,y_pred2)
accuracy
```

```
0.9775280898876404
```

```
[69] from sklearn.metrics import confusion_matrix, accuracy_score, precision_score, recall_score, f1_score
     cm = confusion_matrix(y_test, y_pred2)
     accuracy = accuracy_score(y_test, y_pred2)
     precision = precision_score(y_test, y_pred2)
     recall = recall_score(y_test, y_pred2)
     f1 = f1_score(y_test, y_pred2)
     print("Confusion Matrix:")
     print(cm)
     print("Accuracy:", accuracy)
     print("Precision:", precision)
     print("Recall:", recall)
     print("F1-Score:", f1)
```

```
Confusion Matrix:
[[73  0]
 [ 2 14]]
Accuracy: 0.9775280898876404
Precision: 1.0
Recall: 0.875
F1-Score: 0.9333333333333333
```

PREDICTION ACCURACY

WITH THE GIVEN DATASET, WE ACHIEVED AN IMPRESSIVE ACCURACY OF **97.7%** USING THE **RANDOMFOREST CLASSIFICATION**. THIS HIGH LEVEL OF ACCURACY UNDERSCORES THE MODEL'S EFFECTIVENESS IN PREDICTING EMPLOYEES' PREFERRED TRANSPORT MODES BASED ON THEIR PERSONAL AND PROFESSIONAL ATTRIBUTES. IT DEMONSTRATES THE POTENTIAL FOR **OPTIMIZING TRANSPORTATION SERVICES** AND ENHANCING EMPLOYEE SATISFACTION WITHIN THE ORGANIZATION.

INPUT :

```
input_data=(17,0,1,1,1,1,8,0)
input_np_data=np.asarray(input_data)

input_data_reshaped=input_np_data.reshape(1,-1)

std_data= sc.transform(input_data_reshaped)
print(std_data)

pridictions=classifier.predict(std_data)
print(pridictions)
```

```
[[-2.37963942 -1.58270354  0.55777335  1.75184632 -1.0401606  -1.44511054
  -0.91158721 -0.56212555]]
[0]
```

Thank You

Team :

Sumresh N
RA2111003020402

Abhilash R
RA2111003020423

Moparthi Surya Teja
RA2111003020419

S R Dwaraknath
RA2111003020414