WORKSPACE ZOHO OFFICE

Problem:

The goal of this project is to determine if Zoho is considered a good company for employees based on employee data. This can be framed as a binary classification problem where the model predicts whether an employee views Zoho as a "good" company or not. Using historical employee data, the objective is to train a predictive model that can accurately classify the outcome based on several factors.

Predictive Model

We use a **Decision Tree Classifier** to build a predictive model. The input data consists of several features (columns), except for the target feature, Zoho_Good_Company, which indicates whether employees consider Zoho a good company (1 for good, 0 for not good).

1)Input (Features): All columns in the dataset, except Zoho_Good_Company.
2) Output (Target): Zoho_Good_Company (whether Zoho is considered a good company or not).

Dataset Sample

Workspace_OK: Indicates whether the workspace is satisfactory for the employee (1 for "Yes", 0 for "No").

Internet_OK: Specifies if the internet connectivity at the workplace is adequate (1 for "Yes", 0 for "No").

System_Resource_OK: Indicates if the system resources (hardware, software, etc.) provided to the employee are sufficient (1 for "Yes", 0 for "No").

Team_OK: Represents if the employee is satisfied with their team environment (1 for "Yes", 0 for "No").

Skill_Improvement: Denotes whether the employee feels they are improving their skills at the company (1 for "Yes", 0 for "No").

Salary_OK: Indicates if the employee is satisfied with their salary (1 for "Yes", 0 for "No").

Lunch_OK: Shows whether the employee is happy with the lunch services provided (1 for "Yes", 0 for "No").

Parking_OK: Specifies if the parking facilities meet the employee's expectations (1 for "Yes", 0 for "No").

Restroom_OK: Represents whether the restroom facilities are adequate (1 for "Yes", 0 for "No").

Air_Ventilation_OK: Indicates if the air ventilation in the workplace is satisfactory (1 for "Yes", 0 for "No").

Zoho_Good_Company: The target column that indicates whether the employee considers Zoho a good company overall (1 for "Yes", 0 for "No").

Algorithm

The algorithm used in this project is the **Decision Tree Classifier** from the **sklearn** library. A **Decision Tree** is a **supervised machine learning** algorithm used for both classification and regression tasks. It works by recursively splitting the dataset based on feature values, making decisions at each node until it reaches a leaf, which contains the final prediction.

Why Decision Tree?

1) **Interpretability**: Decision trees are easy to understand and interpret. The model's decision-making process can be visualized, which helps in explaining how the classification is made.
2) **Handling Non-linear Data**: Decision trees can handle complex, non-linear relationships between features and the target variable.
3) **Minimal Data Preparation**: Unlike other algorithms, decision trees don't require much pre-processing, like scaling of data or normalization.
4) **Overfitting Control**: We can fine-tune the tree's depth and other hyperparameters to prevent overfitting.

Why This Algorithm was Chosen Over Others

Decision Tree Classifier was chosen for the following reasons:

1. **Simplicity and Interpretability**: Decision trees provide a clear path from input features to final predictions, making them an ideal choice for binary classification problems like this one.
2. **Versatility**: It can handle both categorical and numerical data effectively, and Zoho's employee dataset might have a mix of such features.
3. **Suitability for Small Datasets**: Decision trees perform well even with relatively smaller datasets, which could be the case here.
4. **Alternative Algorithms**: Other algorithms like logistic regression and support vector machines could also have been considered, but they require more feature engineering and tuning. Decision trees perform well **without much preprocessing**.

Code

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
df = pd.read_csv("zoho_workplace_dataset.csv")
X = df.drop('Zoho_Good_Company', axis=1)
y = df['Zoho_Good_Company']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
model = DecisionTreeClassifier(random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f" Accuracy: {accuracy:.2f}")
if y_pred.mean() > 0.5:
    print("Zoho is a good company!")
else:
    print("Zoho is not a good company.")
```

Output:

```
 Accuracy: 0.80
Zoho is a good company!
```

Conclusion

The decision tree model outputs a prediction on whether Zoho is considered a good company for employees based on the my dataset. The accuracy of the model can be evaluated, and based on the prediction (y_pred), we can conclude whether Zoho is perceived as a good company or not based on employee data.Finally , **Zoho is a Good Company** based on my dataset