

Scaling Techniques Documentation

Introduction:

In the Customer Churn Analysis project, we applied feature scaling to standardize our dataset before performing clustering analysis. Scaling ensures that numerical features are on the same scale, improving the performance of distance-based algorithms like K-Means.

Scaling Technique Used: StandardScaler

We used StandardScaler from sklearn.preprocessing. This method standardizes features by removing the mean and scaling to unit variance. It transforms the data such that the distribution has a mean of 0 and a standard deviation of 1.

This is important because features like MonthlyCharges and Tenure have different ranges, and K-Means is sensitive to the magnitude of data.

Code Snippet:

python

CopyEdit

```
from sklearn.preprocessing import StandardScaler
```

```
scaler = StandardScaler()
```

```
scaled_data = scaler.fit_transform(df[['tenure', 'MonthlyCharges', 'TotalCharges']])
```

The df DataFrame contains the selected features, and scaled_data is the resulting standardized matrix.

Conclusion:

By applying StandardScaler, we ensured that all input features contribute equally to the clustering process. This prevents bias toward features with larger numerical ranges and enhances the interpretability of clusters.