**Feature Scaling and Normalization for Data Preparation**

**Name:**

**ID:**

**Table of Contents**

[1. Introduction 3](#_Toc177147723)

[2. Dataset Description 3](#_Toc177147724)

[3. Scaling Methodology 3](#_Toc177147725)

[4. Scaling Implementation 3](#_Toc177147726)

[5. Results of Scaling 4](#_Toc177147727)

[6. Handling Categorical Variables 5](#_Toc177147728)

[7. Conclusion 5](#_Toc177147729)

# Introduction

Scalings and encodings are the main processes of data preparation step referred to as preprocessing. As for quantitative predictors, they are normalized to attain measure within comparable range Numerical and categorical variables are transformed differently where the former is normalized while the latter is one-hot encoded to binary indicators. Such techniques enhance model performance since normalization makes data appropriate for feeding to species of a machine learning algorithm.

# Dataset Description

The given data set has the quantitative as well as qualitative variables. While comparing analogues of Numerical features mentioned above it was observed that values are present in the scaled down form where values of tenure are small in magnitude in comparison to MonthlyCharges and TotalCharges. However, Contract\_One year, Contract\_Two year and InternetService\_Fiber optic being categories of data requires to be encoded in an integer form to be used for the algorithms by one-hot encoding.

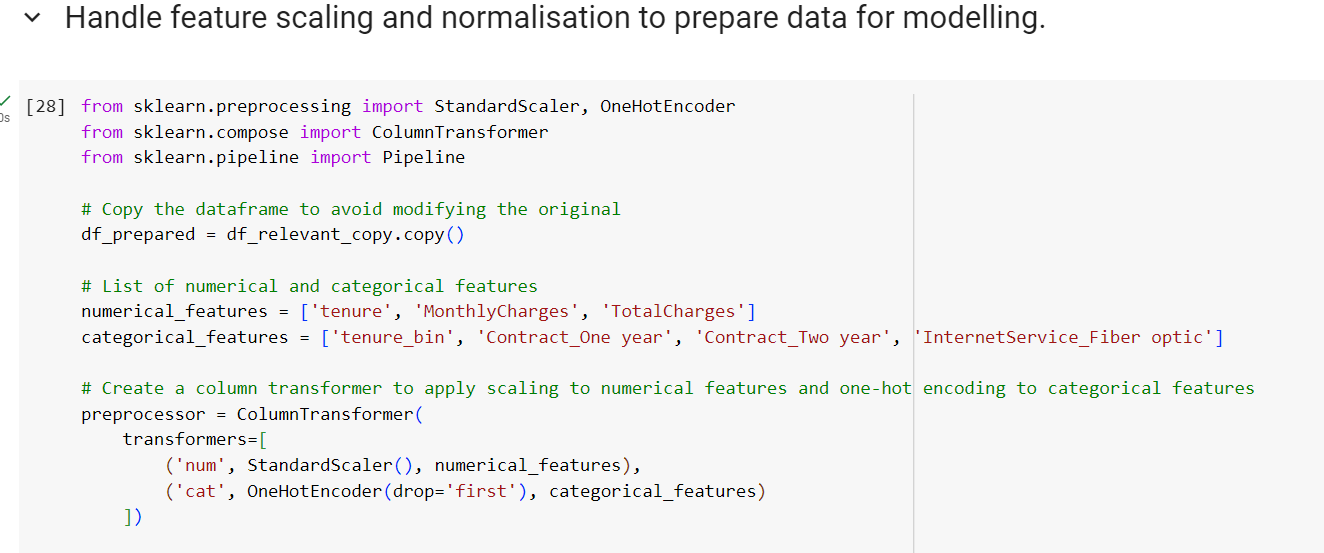
# Scaling Methodology

The StandardScaler from sklearn. Standardization was applied under preprocessing to the continuous variables was applied. Standard scaling is also known as z-score scaling and it subtracts the mean from each feature and brings them to unit variance. This helps make sure that all the features with numeric values are standardized with a mean of 0 and a standard deviation of 1; thus, all features are equally important throughout the model’s learning process.

For categorical features, One hot Encoder was used in converting them to binary columns so that they could be understood by the model.

# Scaling Implementation

The following code was used for feature scaling and normalization and where StandardScaler is used for numerical features and OneHotEncoder for categorical one.



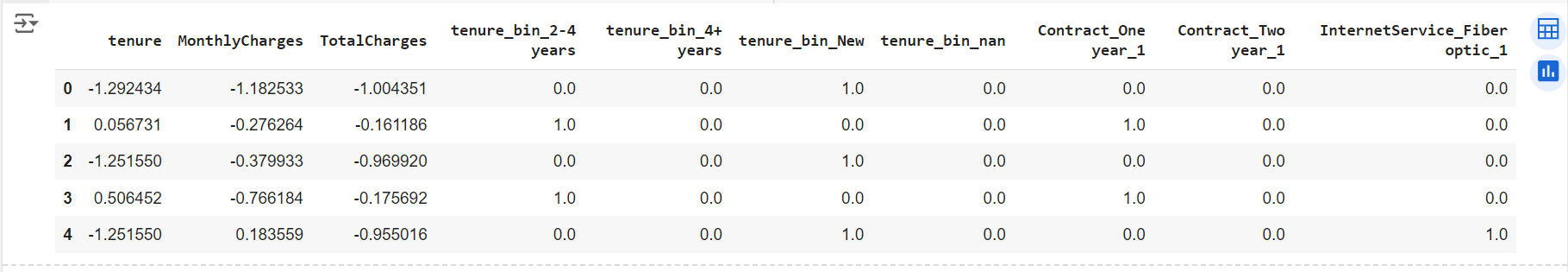
A screenshot of a computer code

Description automatically generated

# Results of Scaling

After applying the StandardScaler, the numerical features (tenure, MonthlyCharges, TotalCharges) were transformed to have a mean of 0 and standard deviation of 1. For example:

* **Before Scaling**: MonthlyCharges ranged from 18.85 to 118.75.
* **After Scaling**: MonthlyCharges has a mean of 0 and standard deviation of 1, ensuring that it no longer dominates the training process due to its scale.



For categorical features, OneHotEncoder created additional binary columns representing each category. For instance, the Contract\_One year column was transformed into a binary column with values of 0 and 1.

# Handling Categorical Variables

Other data types include Contract\_One year that was encoded by OneHotEncoder along with other categorical data types including Contract\_Two year and InternetService\_Fiber optic. This method develops dummy variables for each of the categories of the original categorical variables. The new binary columns are named based on the original feature name and the category given thus making it easier to interpret the results.

# Conclusion

To increase the accuracy of model and preparing the data to be understandable by the machine learning model, it was crucial to scale the numerical features and encode the categorical features. StandardScaler served to normalize the numerical features while OneHotEncoder was of aid in transforming the categorical features into a format that was more consumable for the model training procedure.

I was able to pest the dataset through these techniques in a manner that would enhance the model’s readiness throughout the modeling process.