Netflix movies and tv shows analysis report

1. Introduction

Project Overview

The goal of this project is to classify various shows into movies or tv shows category.

Dataset Description

The dataset contains 8807 tweets with the 12 columns (became 15 after feature engineering):

2. Data Preprocessing

Data Cleaning

• autoclean() and klib.datacleaning() was not used

• Removed unnecessary columns such as shows_id

• Handled missing values by filling them as appropriate (mode, median).

• Converted the type column to numerical values using mapping as it is

target value.

• encoding was done manually

• scaling was done only on features assigning them to x variable and removed the target

value temporarily to prevent it from scaling and becoming continuous value

3. Model Training

Splitted the dataset into training and test sets using an 80-20 split.

Model Selection

one machine learning model was considered:

• LogisticRegression

4. Model Evaluation

The following metrics were used:

• Classification report

• Accuracy

• Presicion

• F1 score

Results

	precision	recall	f1-score	support
0	0.76	0.67	0.72	282
1	0.85	0.90	0.88	599
accuracy			0.83	881
macro avg	0.81	0.79	0.80	881
weighted avg	0.83	0.83	0.83	881