ML for prediction of viewership metrics for the streaming company using the Time Series approach.

Preface

The company X operates in the video streaming entertainment industry. It owns a huge portfolio of different brands, the brands in turn consist of several seasons, the seasons are divided into series.

All content streaming happens on its own platform, where users choose what to watch and when to watch it.

All data on views and brand subscriptions are collected in a regular mode and stored inside the Google's BigQuery because company's infrastructure build on Google Cloud Platform (GCP)

Problem

Since it makes no sense to stream content endlessly, stakeholders would like to know when interest in a particular brand starts to drop.

In addition, they would like to know in advance the approximate performance of streaming a given brand (number of views, popularity).

Scope and metrics

As the problem stated in Time-Series analysis and prediction domain the following metrics were used to decide the success of the solution:

Offline Metrics:

- Mean Absolute Error (**MAE**)

  Measures the average magnitude of errors in a set of predictions without considering their direction or how close the forecasts are to the actual values

- Mean Absolute Percentage Error (**MAPE**)

  Expresses accuracy as a percentage which can be easier to interpret for stakeholders and non-technical persons on the model's performance

Online Metrics:

- Viewer Engagement

  Metrics like average watch time, session duration and frequency of visits will indicate how engaging the content is

- Content Performance

  Metrics like the number of new subscribers attributed to a particular brand show or the increase in viewership after promoting certain content

Dev Process


Problem Statement
Data Collection and Data Engineering
Research and Exploration Analysis
Model Research and Training
Evaluation and Deployment
Integration and Monitoring
Feedback Loop

Solution Architecture