## **1.2 Project goals**

The overall goal of this paper is to develop a prediction module for choosing ideal time to broadcast emission TV, which utilizes supervised machine learning to help automate the process of finding the best time of broadcast in the daily TV-guide.

Specifically, this project

• evaluate the prediction accuracy of two or more machine learning models to find the best model for predicting TV-guide,

• train and tune the best model to give a prediction accuracy error

# **3. METHOD**

In this chapter, the methods, and the arguments why these methods were used in this paper will be presented. A literature review was conducted to find and determine the types of machine learning models that could be used for the problem. The study included previous studies, research, and work about supervised machine learning. We then evaluated the selected models and compared them using different methods. Finally, the evaluation and comparison results were collected and analyzed.

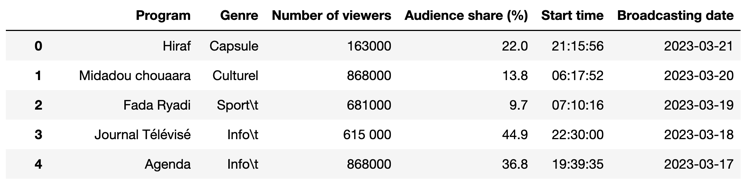
This work uses data from two years of daily top ratings that represent the success of television programs of various genres: Capsule, Cultural, Sport, News, Music, documentaries, Religious, Political and series. It is also divided into: Program Name, Genre, Number of viewers, Audience share (%), Start time and Broadcasting date.

**4. DATA**

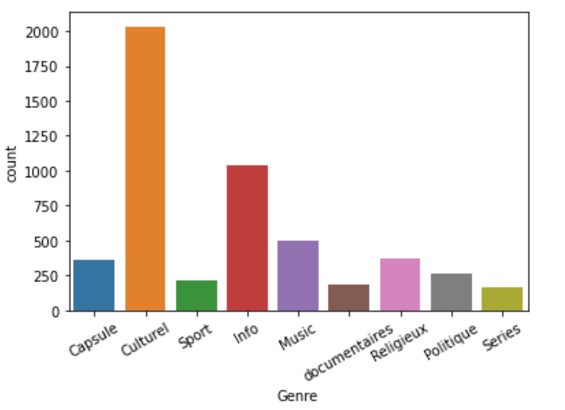
This chapter introduces the data used during model training and evaluation, and the process of formatting the collected raw data.

**4.2 Getting data Raw**

Since it is not possible to measure the true television viewing rate, the published TV rating values are used, which are estimated by CIAUMED, a company that, among other things, measures TV ratings in Morocco. The data obtained was historical data from 2021 at the time of writing this paper with the features (columns) shown in “Figure.1”, and Divided into 9 categories “Figure.2”. This data is also calculated from the control room of the Moroccan television station (SNRT) by high-performance equipment, “Figure.3” and “Figure.4”, to calculate the measurement of satellite television broadcasts.

****

**Figure 1:** . Example of data from CIAUMED

****

**Figure 2:** . data categories

**3.2.2 Data features and formatting**

The acquired data set contains many variables, and in our study we need:

**Program**: The name of the TV program

**Genre**: Genre of TV show

**Number of Viewers**: The number of viewers for each TV program

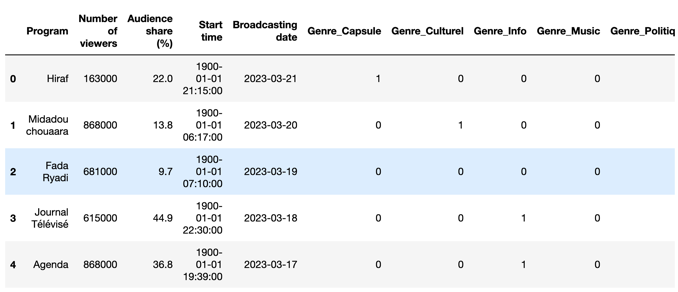
**Audience Rate (%)**: The viewership rating compared to other TV programs on the other Moroccan TV station

**Start time**: The start time of the broadcast

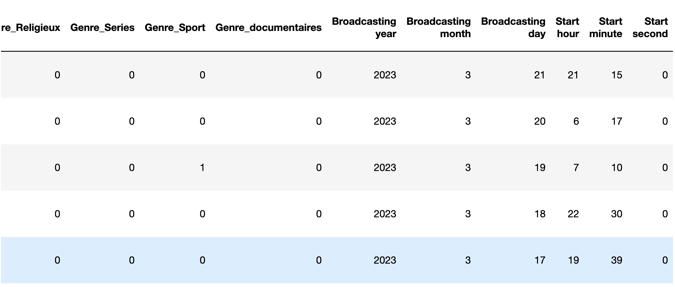
**Broadcast Date**: The date of program broadcasting

**3.2.2.8 Formatted data**

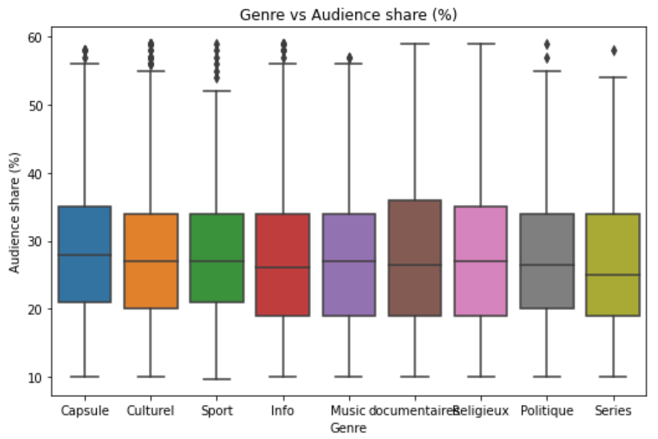
In the datasets is the time a program starts during a day, as shown in “Figure.5” and “Figure.6”, are two halves of the same table, divided for illustration purpose. Changes made to the data was the removal of “:”. It was removed due to the models interpreting this as "/". to convert to a numeric value instead of text, and the same for dividing the dataset into TV program categories “Figure.7”.

****

**Figure 5:** Example of data used for prediction models first half

****

**Figure 6:** Example of data used for prediction models second half



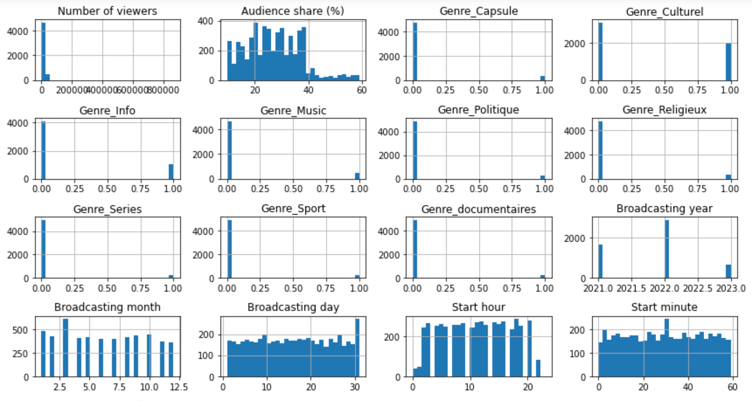
**Figure 7:** . Plot Visualization for rating of Genre of TV Programs

**3.3 Evaluation of the results**

This chapter describes the process of evaluating results. We compared the models to see which one performed the best. We then compared the best models with experts to compare TV prediction accuracy.

**3.3.2 Training process**

After Checking distribution of data “Figure.8”, a training split of 80/20 was used in the training process. 80% of the data in the dataset was used for training and 20% of the data in the dataset was used for testing after training the model. The 80/20 split was chosen after showing the best results from the splits tested. we ran the training many times for each model to see if there was any difference in the prediction results.



**Figure 8:** . Plot Visualization for Checking distribution of data

**3.3.3 Testing process**

The focus was on prediction accuracy to see how well the models compare. The model with the best predictive accuracy was selected for further analysis. The best performing model among other models was evaluated using the mean squared error.

**4 Results**

This chapter will present the results of fitting the hyperparameters of the models, the performance of the models against each other, and the performance of the best model against the data set. As well as the development of a simple web page in a Python flask with our model in a Pickle file, “Figure.9”, to make it easier to use for the prediction team in the television channel.

The best performance scores are for Random Forest and K-nearest Neighbors Regression “Table.1”.

**Table 1:** Score for the best algorithm performance

|  |  |  |  |
| --- | --- | --- | --- |
| Model | RMSE | MAE | R2 Score |
| Ridge | 2.6974 | 2.1626 | 1 |
| LR | 2.37430 | 2.05145 | 1 |
| KNN | 1.243153 | 7.58540 | 1 |
| DT | 2.77836 | 1.10428 | 1 |
| RF | 1.02544 | 3.79929 | 1 |

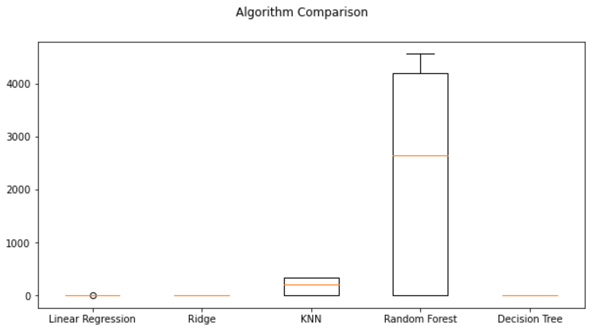
The training and testing were both done on pseudorandom subsets of the dataset. To make sure the training and testing were thorough, each model was trained and tested many times and then compared against each other.

**4.2.3 Comparison of the machine learning models**

The following were the results after training the KNN model ten times. MAE had a maximum value of 7,58540, minimum value of 9,6760. RMSE had a maximum value of 1,24, minimum value of 5,96.

The results of the models compared to mean values with standard deviation are presented side by side in “Figure.10”.

As shown, K-nearest Neighbors Regression performed better on both metrics, hence K-nearest Neighbors Regression, was chosen for comparison with current processes of estimating the best TV show to have more views by choosing airing by date and time.



**Figure 9:** . Extract Result in a System Report