**Disclaimer**

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying is a grave and serious offence in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion, or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references.

I have not copied or paraphrased an extract of any length from any source without identifying the source and using quotation marks as appropriate. Any images, audio recordings, video or other materials have likewise been originated and produced by me or are fully acknowledged and identified.

This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study. I have read and understood the referencing guidelines found atDCU Academic Integrity and Plagiarism Policy.

I understand that I may be required to discuss with the module lecturer/s the contents of this submission.

I incorporates us in the case of group work, which is signed by all of us.

Name: Rasika Mehta

Date: 27/04/2021

# Predicting Video Memorability using Linear Regression, Decision Tree Regressor, and Random Forest Regressor

Rasika Mehta
MSc. Computing - 2020-21
Dublin City University, Dublin, Ireland
Course: MCM1 - Data Analytics
20210036

***Abstract*— The purpose of predicting video memorability is to predict the quality of a video based on various features. People are presented with numerous photos and videos consistently; however, they have this extraordinary capacity to recall the subtleties, despite the fact that a considerable lot of them look basically the same. It is important so that we can organize and retrieve digital content and use it as per our needs. Semantic features and features like captions are useful in finding video memorability. In this, features like InceptionV3 and C3D with linear regression and decision tree regressor machine learning models are used. This prediction is made using 2000 test features of each type. The dataset consists of 6000 dev- set videos and 2000 test-set videos. It comes with long-term and short-term annotations. Video Memorability is predicted using decision tree regressor as it gives better output with Spearman's Ranking coefficient.**

## I. INTRODUCTION

In this work, I research the utilization of different visual and semantic features to foresee video memorability and direct a broad examination over the picked features, to build up a good video memorability prediction model. Features like HMP, LBP, ColorHistogram, C3D, InceptionV3, Aesthetic features, etc., are provided for 6000 dev videos and 2000 test set videos. In this, C3D(Convolution 3D) feature and InceptionV3 feature are selected. InceptionV3 is a convolutional neural network model used for video classification. The two features are extracted and saved in a .npy file and trained on Linear Regression and Decision Tree models. Evaluation of models is done using Spearman's Ranking Coefficient. Here, C3D is used as it is a video feature that consists of images, objects, scenes, and other categories.[5] InceptionV3 is used for video classification and can be a significant feature in predicting video memorability.

### *A. key Findings:*

1) Long Term scores are better than short term scores with each model.
2) C3D with linear random forest gives lower performance than other models.
3) The best performing model is the Decision tree with both the features combined.

## II. RELATED WORK

Researches on video memorability center around anticipating the likelihood of the substance being recalled, which bodes well in both academic types of examination and reasonable business.[2] In the study by Raghavendra Pratap Singh, shows that multiple features of videos and multiple models like regression, ensemble, etc can be used to compare the scores and find the best model for predicting video memorability.[1]

A research states that "very little work has been done related to machine learning and computational science on video memorability."[2] Psychological perspective shows that people store visual features in long-term memory. In studies done by Gaurav Goswami, Romil Bhardwaj, Richa Singh, Mayank Vatsa, They have used Augmented Deep Learning model for video face memorability. [3] The proposed algorithm, named as MDLFace, is assessed on two openly accessible video face data sets, Youtube Faces and Point and Shoot Challenge.

## III. FEATURE EXTRACTION

To predict video memorability, I have used the C3D feature and InceptionV3 features. C3D is a video feature that provides objects, images, and scenes of a video. By training a large deep learning dataset, we can obtain C3D features.[5] The test set of 2000 videos were given with precomputed C3D and InceptionV3 features each. InceptionV3 is a CNN model used for image recognition previously trained on the IMAGENET dataset.[7]

Dev-set features for training machine learning models were extracted along with Test-Set features. After extraction, features were loaded and saved in .npy files. There were three data frames in InceptionV3 feature files, namely 0,56, and 112. First, I extracted all features from 3 data frames, and then they were appended in a single file forming a total of 6000 features in Dev-Set and 2000 in Test-Set.

## IV. MODEL TRAINING AND IMPLEMENTATION

Here three machine learning models are used to train and calculate Spearman's Ranking Coefficient. Comparison is made between the three using scores obtained after implementing them with extracted Features of videos. The models used are as follows:

1)Linear Regression
2)Decision Tree Regressor
3)Random Forest Regressor

After calculating the score, it is clear that decision tree regressor algorithm with combined C3D and InceptionV3 features outperforms the other two models.

Decision tree algorithm can be performed on non-linear data, whereas linear regression supports linear data. with less noisy data and a large number of features, linear regression can outperform decision tree and random forest.[6] Decision tree does not need rigorous training of data in comparison with random forest.[8] It is comparatively faster than random forest and is easy.

## V. RESULTS AND DISCUSSION

After training and fitting the given dataset into machine learning models, Spearman' Ranking Coefficient was calculated using each model on C3D features and InceptionV3 features as well as combined features. They are as shown in the table below:

| | | C3D | InceptionV3 | C3D + InceptionV3 |
|---|---|---|---|---|
| **Linear Regression** | *Short Term* | 0.015 | 0.011 | -0.012 |
| | *Long Term* | -0.008 | 0.016 | 0.004 |
| **Decision Tree** | *Short Term* | 0.03 | 0.001 | **0.052** |
| | *Long Term* | 0.027 | 0.03 | **0.042** |
| **Random Forest** | *Short Term* | 0.027 | -0.069 | -0.026 |
| | *Long Term* | 0.039 | 0.008 | 0.025 |

Table 1: Short Term and Long Term Memorability Scores Form Multiple Machine Learning Models

As we can see from the above table, the Decision tree regressor gives the best score among the three models. It is chosen to be implemented on test features to predict video memorability. It takes lesser computing time than random forest. All the results after calculating scores for all the videos are stored in Results CSV. There we can observe that the decision tree model is giving better performance.

## VI. CONCLUSION

The short-term and long-term memorability scores are not totally dependent on C3D and InceptionV3 features. But we can implement various models like linear regression, decision tree and random forest to compare and study which model best suits for predicting video memorability. In this, the Decision tree comes as outperforming model for calculating scores. In the future, features like HMP, Aesthetic features, video captions can be taken into consideration for prediction purposes. Also, Ensemble models and neural networks can be implemented on video features.

## REFERENCES


[1]Raghvendra Pratap Singh in "Predicting Video Memorability" *depicted an analysis of pre-extracted features like captions, C3D, HMP etc. with the prediction of short- and long-term memorability using assorted machine learning algorithms.*

[2]A. Kar, P. Mavin, Y. Ghaturle and M. Vani, "What Makes a Video Memorable?," *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2017, pp. 373-381, doi: 10.1109/DSAA.2017.37.*

[3]G. Goswami, R. Bhardwaj, R. Singh and M. Vatsa, "MDLFace: Memorability augmented deep learning for video face recognition,"*IEEE International Joint Conference on Biometrics, 2014, pp. 1-7, doi: 10.1109/BTAS.2014.6996299.*

[4] Yue F., Li J., Sun J. (2021) Insights of Feature Fusion for Video Memorability Prediction.*In: Zhai G., Zhou J., Yang H., An P., Yang X. (eds) Digital TV and Wireless Multimedia Communication. IFTC 2020*. Communications in Computer and Information Science, vol 1390. Springer, Singapore. https://doi.org/10.1007/978-981-16-1194-0

*[5]* Du Tran, Manohar Paluri "C3D: Generic Features for Video Analysis" on *December 8, 2014*

[6] "Comparative Study on Classic Machine learning Algorithms" *by Danny Varghese on December 6, 2028*

[7] *Researches at Apriorit "Using Modified Inception V3 CNN for Video Processing and Video ClassificationIt was originally published on https://www.apriorit.com/"* on March 28, 2019

*[8] "Random Forest Vs Decision Tree: Difference Between Random Forest and Decision Tree" by Pavan Vadapalli on December 30, 2020*