

The Anatomy of Successful YouTube Content: A Metadata-Based Study of Factors Contributing to Channel and Video Success



Manan Patel and Jake Shoffner



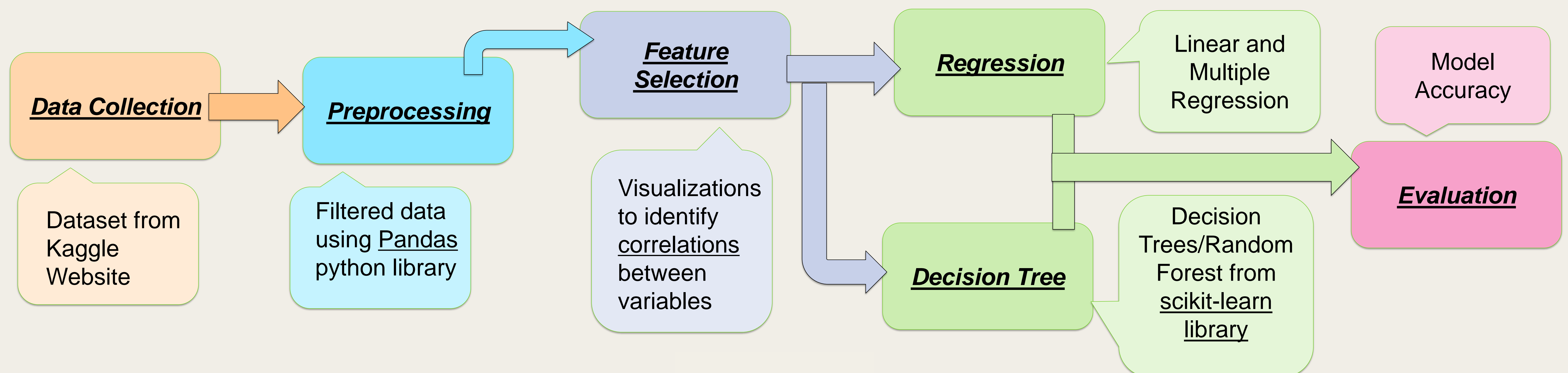
Motivation and Goals

- We assist individuals seeking success on the YouTube platform by identifying the key variables that determine a video and channel's success.
- We utilize **predictive modeling techniques**, such as **linear regression** and **decision trees** to identify the critical variables that play the most significant roles in determining success in channel and videos.

Dataset

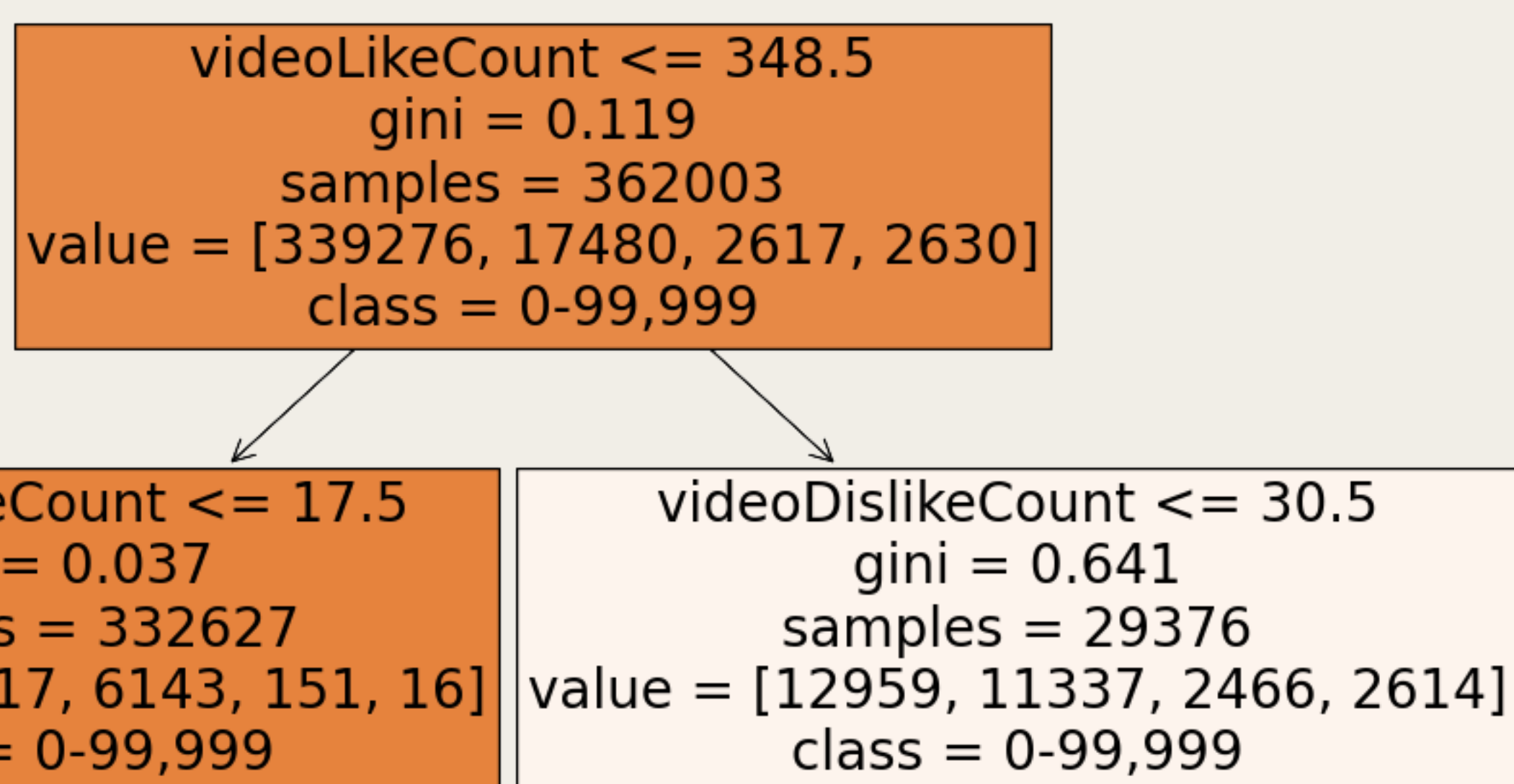
- The **YouTube Videos and Channels Metadata** dataset includes entries for over 500,000 videos uploaded to the YouTube platform.
- Dataset obtained from the **Kaggle** website [1].
- Contains **575610 rows** and **26 columns**.
- Sample Entry: < *videoViewCount*, *subscriberCount*, *videoLikeCount*, ... >

Methodologies

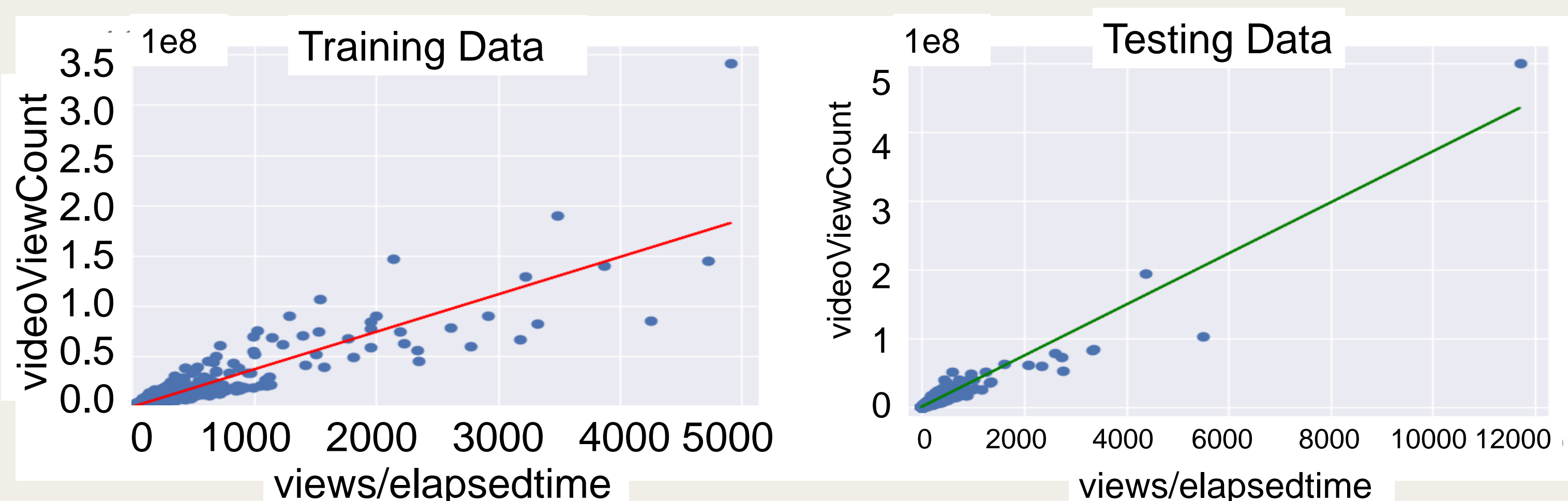


Predicting Video View Counts

- Decision Tree:** 4 range classifications, **depth 4** (95.54% accuracy)
- Decision Tree:** 4 range classifications, **full depth** (98.39% accuracy)



- Linear Regression:** Predicted video view counts based on views to elapsed time ratio (91.47% model accuracy)



Conclusion

- Using a classification model decision tree increases accuracy of predictions at the cost of being less specific than a regression model decision tree.
- The five most important predictor variables in our decision tree model are videoLikeCount, videoDislikeCount, views/subs, subscriberCount, and channelViewCount.
- The best predictor variable in our linear regression model is views/elapsedtime.

Predicting Channel Subscriber Counts

Multiple Linear Regression Model:

78.86% model Accuracy

We input variety of variables into a multiple linear regression model to identify the key variables in predicting How many subscribers a channel has.



Conclusion

The best predictor variables are:

- channelViewCount*
- subscriberCount*
- videoViewCount*
- videoLikeCount*.

