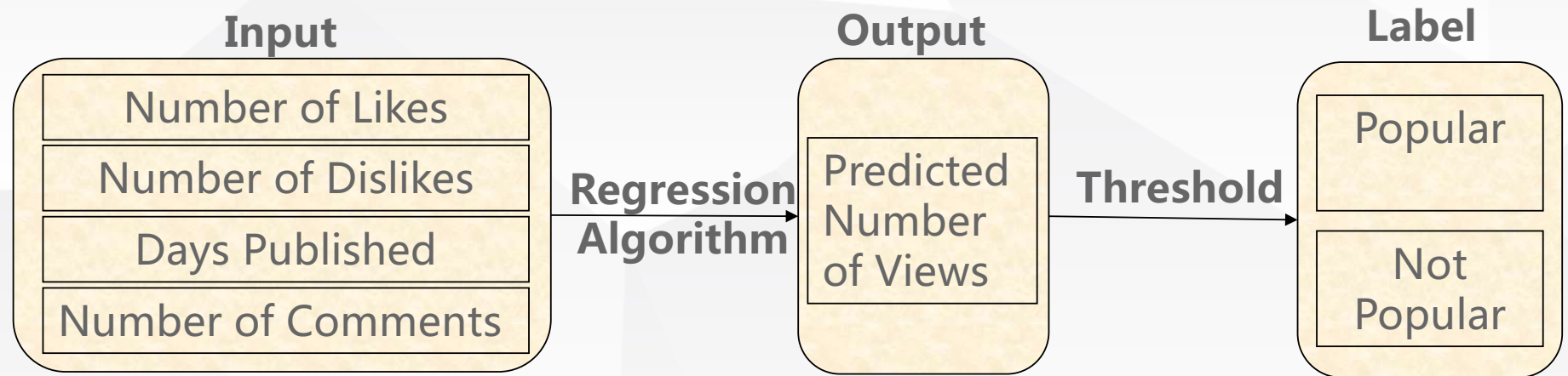


Data Preprocessing

Model Architecture



Dataset Selected

375942 pieces of YouTube data from 10 countries

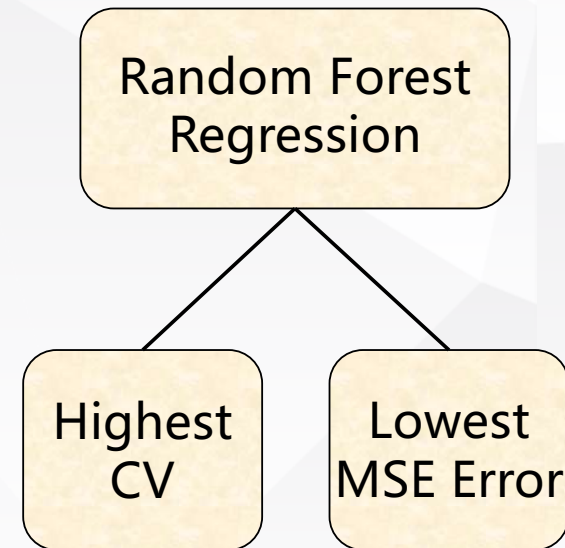


Algorithm Selection

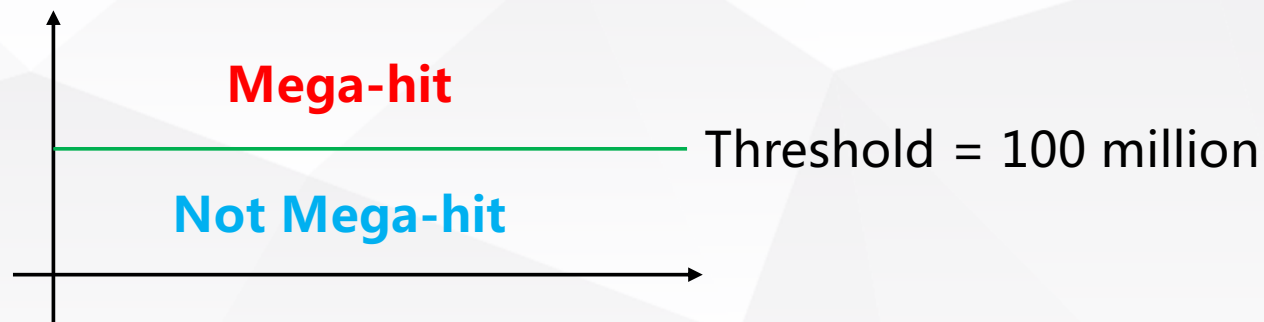
Regression Algorithm Selection

Algorithm	CV	MSE Error
Linear Regression	0.6953	15,433,411,971,867.74
Ridge Regression	0.6953	15,433,411,971,866.65
Random Forest Regression	0.9703	1,084,290,013,297.92

Selected Regression Algorithm



Threshold Selection

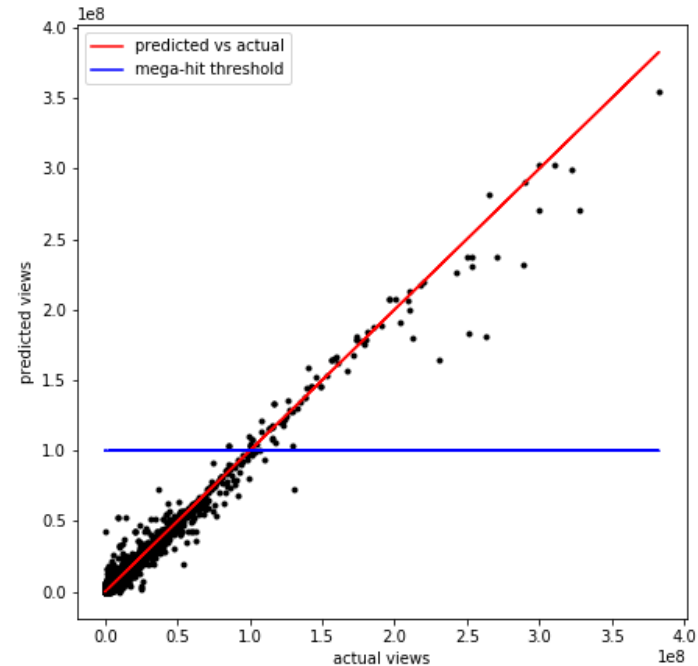


Analyze and Conclusion

Result Analyze

Parameter	Value
Precision	0.935484
Recall	0.966667
F1-score	0.950820

Most points locates on or near **red line**, where predicted views original number of views; Even if the difference is large, it usually does not affect the relative position with **blue line**.



Conclusion

The developed model could well predict the **number of views** of a video and whether it is **popular or not**, given the **number of likes, dislikes and comments**, and **days it has been published**.