

Project Initialization and Planning Phase

Date	June 2024
Team ID	7740096
Project Name	The Language Of Youtube: A Text Classification Approach To Video Descriptions
Maximum Marks	3 Marks

Define Problem Statements(Customer Problem Statement Template):

The business problem in this scenario is to help clients identify the appropriate category of YouTube videos to advertise their product based on the text descriptions of the videos. This is a challenging task since YouTube has a vast collection of videos covering different topics and categories, and it can be difficult for clients to identify which category of videos will be the most effective for advertising their product. By developing a text classification model based on NLP and deploying it using Flask, this project aims to provide clients with a tool to make more informed decisions about where to advertise their products on YouTube. Reference: <https://github.com/Kamya-Paliwal/The-Languageof-YouTube-A-Text-Classification-Approach-to-VideoDescriptions/blob/main/app.py> **Example:**

Customer Problem Statement

I am	CUSTOMER
I'm trying to	Predict the appropriate category of YouTube videos to advertise their product based on the text descriptions of the videos.
but	The Estimation levels are Not Accurate
because	The System Components Will Not Work Until It Has Some descriptions with very short or informative content
which makes me feel	anxious and dissatisfied

Problem Statement (PS)	I am	I'm trying to	But	Because	Which makes me feel

PS	A data scientist working on The Language Of Youtube: A Text Classification Approach To Video Descriptions	Develop a predictive model for to predict the best video category for businesses to advertise their products on YouTube	I'm encountering difficulties with data quality and the complexity of Video descriptions can be in various languages, making it challenging to process them uniformly.	the available data is often incomplete, inconsistent, and influenced by numerous unpredictable factors such as Incomplete Descriptions, Inconsistent Formatting	concerned about the reliability and accuracy of the predictions, impacting the effectiveness on the Presence of spammy, irrelevant, or promotional content
----	---	---	--	---	--