MOTIVATIONAL YOUTUBE VIDEO CLASSIFICATION AND PERFORMANCE COMPARISION BASED ON TITLE

Table of Contents

Abstract 1
Introduction 2
Literature Survey 4
Problem Statement 11
Objective 12
Dataset 13
Model Evaluation 15
Proposed Method 22
Evaluation Parameter 33
Result & Analysis 36
Conclusion 43
References 44

Abstract

YouTube has been the leading online video service for many years, attracting billions of subscribers and viewers. Among the wide range of content available on YouTube, there are numerous channels dedicated to providing motivational videos. Motivation stems from the word 'motive,' which refers to individuals' needs, desires, wants, or drives. On YouTube, motivational videos can be categorized into two types: extrinsic motivation and intrinsic motivation. Extrinsic motivation is driven by external factors and often involves rewards from outside the individual, such as trophies, money, social recognition, or praise. On the other hand, intrinsic motivation is internal and originates within the individual, where the satisfaction of solving a complex crossword puzzle serves as its own reward. To address the challenge of distinguishing between extrinsic and intrinsic motivational videos on YouTube, we propose an analysis approach based on natural language processing (NLP). This approach enables the classification of videos into their respective categories. Once the videos are classified, we perform a performance comparison by analyzing the titles using various common machine learning methods such as Support Vector Machine, Naive Bayes, Logistic Regression, Random Forest, etc. Additionally, we utilize deep learning classifiers such as Convolutional Neural Networks, Long Short-Term Memory, Recurrent Neural Networks, etc., to further enhance the performance analysis.

Introduction

In today's society, with the rapid advancement of technology, various social media platforms are constantly competing to attract users. Among these platforms, YouTube stands out as one of the most popular and rapidly growing platforms. As of 2019, according to Alexa, YouTube ranked as the second top website globally [1]. The widespread popularity of YouTube can be attributed to its convenience, allowing users to watch videos anytime and anywhere, and its accessibility to people from all walks of life. This has contributed to YouTube's continuous growth as an industry.

YouTube is an online video platform where various types of videos are uploaded. YouTube is the second largest search engine next to Google. People upload more than 100 hours of video per minute to YouTube. It's one of the best ways to communicate with a wide audience. Nowadays people of all ages are viewers of YouTube. It is an extremely important platform for individuals. There are different varieties of channels are entertainment, news report, food, gaming, beauty and fashion, music, sports, Science and Technology, travel, etc.

And now when the topic is Motivational Video, then undoubtedly YouTube is the best platform for it. On YouTube to attract audiences, the YouTuber uses the most interesting words in the video's title. On YouTube, there are two types of videos: Intrinsic and Extrinsic.

YouTube offers several options for video classification, and various techniques can be employed to classify videos based on their content, comments, and other metrics. However, these methods can often be intricate and time-consuming. Our approach to video classification on YouTube focuses on text classification, with a particular emphasis on video titles. By employing text classification techniques, we aim to categorize videos based on their titles efficiently and effectively. This approach allows us to quickly identify the nature, theme, or purpose of a video by examining the words used in its title.

In summary, our classification approach for YouTube videos focuses on text analysis, specifically targeting the titles. By leveraging text classification techniques, we can swiftly categorize videos based on their titles, enabling us to gain insights into the content and attractively appeal to the audience.

In the 21st century, motivational quotes and videos have become an essential element for many individuals in society, particularly among the youth.

However, it is worth questioning the necessity of motivation and the underlying logic behind the depth of motivational videos. There are instances when we all require a bit of motivation, something to uplift us and reengage with the world. Nevertheless, it is important to acknowledge that motivational videos may not be truly beneficial unless our intention is simply to feel good momentarily. In such cases, we should recognize that we watch these videos for entertainment purposes rather than expecting them to have life-altering effects.

This paper addresses the problem of classifying motivational video titles on YouTube based on their type of motivation. The main focus is to gather YouTube title texts and descriptions for this classification task. Collecting, processing, and classifying text data from YouTube presents significant challenges for researchers. Fortunately, deep learning and natural language processing [2] offer effective solutions for text analysis and processing. For this study, we manually processed and labeled the data from YouTube video titles, classifying them into Intrinsic and Extrinsic sentiments. Several measures and steps are involved in the text classification process.

In this project, we harness the power of deep learning techniques to address complex challenges and achieve groundbreaking results. Deep learning, a subset of machine learning, empowers us to create intelligent systems capable of learning and making decisions from vast amounts of data. By utilizing deep neural networks with multiple layers of interconnected nodes, we can extract high-level features and patterns from raw data. This enables us to build sophisticated models that excel in tasks such as natural language processing and data analysis. Through deep learning algorithms, we can unlock valuable insights, make precise predictions, and drive innovative solutions in our project.

Literature Survey

As YouTube's popularity continues to soar, the field of YouTube text classification is expected to witness further progress, providing deeper insights into user preferences, behavior, and content categorization on this influential platform.

Islam et al. [3] In their study, the researchers performed a performance comparison of several supervised learning algorithms for classifying exaggerated Bangla titles on YouTube. They evaluated a total of six models in their research. Based on the results obtained, the Convolutional Neural Network (CNN) proved to be highly effective in classifying exaggerated Bangla title videos, achieving impressive accuracy, precision, and F1 Score values of 80%, 81.25%, and 76.47%, respectively. These findings demonstrate CNN's superior performance in accurately categorizing YouTube videos with exaggerated Bangla titles, highlighting its potential as a powerful tool for such classification tasks.

Karla et al. [4] researched to conduct a classification of YouTube videos based on their titles and description texts. For this task, they employed a Random Forest Classifier in conjunction with Natural Language Processing techniques such as Bag of Words and Word Stemming. These methods allowed them to effectively process and analyze the textual data to make accurate classifications.

Ni et al. [5] undertook the classification of a large-scale video event on YouTube to address the challenge of general event classification from uncontrolled YouTube videos. As a result, they identified a total of 29,163 video event categories. To achieve this, they utilized a POS-based query method applied to video titles, which enabled them to obtain a substantial dataset of 6,538,319 video samples representing these identified categories. This comprehensive study offers valuable insights into the process of classifying video events on YouTube, highlighting the potential for further advancements in the field of event classification from uncontrolled video data.

Porreca et al. [6] through Text Mining and Sentiment Analysis, an examination of YouTube Italian videos regarding vaccination was conducted. The sentiment analysis revealed intriguing findings. The intensified vaccination campaign, notably promoted by medical professionals, played a significant role in shifting the prevailing sentiment from predominantly negative in 2017 (52% negative) to predominantly positive in 2018 (54% positive). This shift in sentiment indicates the profound impact of the vaccination campaign and the medical community's efforts in influencing public opinion and fostering a more positive outlook towards vaccination in the Italian YouTube community.

Cunha et al. [7] proposed to categorize YouTube comments into three groups: neutral, positive, and negative. The researchers utilized a deep learning-based neural network for this task, which yielded impressive results with an accuracy of up to 84%. This approach demonstrates the effectiveness of deep learning techniques in accurately classifying sentiments expressed in YouTube comments, providing valuable insights for sentiment analysis in social media platforms.

Asghar et al. [8] conducted a comprehensive study, encompassing various aspects of sentiment analysis, event classification, negation detection, social media-aware phrase detection, and prediction of YouTube comments. The main focus of this research was to determine the polarity of the comments, classifying them as positive, negative, or neutral.

Uryupina et al. [9] compiled a dataset by collecting user comments from YouTube videos, and each comment was annotated for both information quality and sentiment polarity. The dataset comprises comments from two types of videos: technical reviews and commercials. By analyzing and annotating these comments, the researchers aimed to understand the information quality and sentiment expressed by users in response to these specific types of videos. This dataset serves as a valuable resource for further research in sentiment analysis, user feedback evaluation, and understanding the dynamics of user engagement with technical reviews and commercials on YouTube.

Bhuiyan et al. [10] employed natural language processing (NLP) techniques to filter useful videos by conducting sentiment analysis on user comments from YouTube. Through their experimental method, they achieved a notable maximum accuracy of 75.435%.

Novendri et al. [11] conducted a sentiment analysis on comments posted on YouTube movie trailers. The comments were categorized into three classes: Positive, Negative, and Neutral. Their dataset consisted of 998 comments. In their study, they evaluated different classifiers, and the Naive Bayes classifier achieved the highest accuracy, reaching an impressive 81%.

Krishna et al. [12] Machine learning techniques were utilized to perform sentiment analysis on YouTube comments related to relevant topics. The researchers aimed to assess whether the trends, seasonality, and predictions derived from the collected videos offer a comprehensive understanding of the impact of real-world events on viewers' sentiments.

Muhammad et al. [13] The researchers employed the Naïve Bayes approach for sentiment analysis, focusing on classifying positive and negative YouTube comments. To further enhance the accuracy and performance, they combined the Naïve Bayes with Support Vector Machine (SVM) using a 7:3 scale, where 70% of the data served as training data, and the remaining 30% was used for testing. This combination resulted in the highest performance test values, including an impressive precision of 91%, a recall of 83%, and an F1 Score of 87%.

Singh et al. [14] conducted extensive research on sentiment analysis, where the analysis of frequently used words has been employed to gain insights into people's emotions. Notably, the analysis focused on the sentiments surrounding the inauguration of Joe Biden as the new President of the United States.

Lavanya et al. [15] In their research, scholars employed deep learning techniques to explore text classification using Natural Language Processing (NLP) in social healthcare networks. The primary objective of this study was to gain valuable insights into training data and classify text by analyzing and extracting raw input through the use of NLP. By harnessing the power of deep learning and NLP, they aimed to develop effective models that could accurately categorize and process textual data within the context of social healthcare networks. The study's findings provide valuable guidance for the implementation of NLP-based approaches in this domain, enabling more efficient and accurate text classification and analysis for improved healthcare communication and decision-making.

Sharmin et al. [16] introduced an attention-based convolutional neural network (CNN) for sentiment analysis on YouTube Bangla text. They curated a dataset comprising 2979 reviews and comments, sourced from social media platforms. These texts were categorized into three sentiment classes: positive, negative, and neutral. By leveraging their attention-based CNN model, the researchers achieved an impressive accuracy of 70% on the validation set.

Aggarwal et al. [17] The researchers presented a novel approach to detect privacy-invading harassment and misdemeanor videos by mining YouTube metadata. To validate their hypothesis, they conducted a series of experiments using an evaluation dataset obtained from YouTube. The empirical results showcased the effectiveness of their approach, as it achieved an accuracy of more than 80%.

Song et al. [18] The researchers utilized Open Metadata to identify video titles in their study. They collected a dataset of 13,291 videos from the actual video-streaming environment of YouTube and conducted their experiments using VTIM. The results demonstrated that VTIM could accurately identify video titles with a remarkable accuracy of 100%. Notably, VTIM achieved this accuracy at a significantly faster rate, nearly 30 times faster than existing methods based on machine learning techniques.

Oliver L et al. [19] In this paper, a novel concept called Interactional Motivation (IM) is introduced, aiming to implement self-motivation in artificial systems. IM differs from extrinsic motivation in that it defines the agent's motivation independently of the environment's state. Moreover, IM is distinct from intrinsic motivation as it allows explicit specification of the agent's inborn value system.

Konstanin A. Grebenyuk et al. [20] In the current research, an empirical model is proposed to cultivate and sustain students' intrinsic motivation for learning. Unlike numerous other motivation models, this particular model does not rely on psychological theories; instead, it stems directly from real-life observations made by experienced learners and educators.

Ying Dong et al. [21] Eco-innovation (EI) aims to achieve dual positive outcomes, benefiting both the environment and the economy. These outcomes encompass enhanced competitive performance and improved environmental performance, representing a double externality characteristic with multiple motivations. Through empirical analyses of 246 enterprises in China, econometric estimations were conducted to assess the impact of various motivations on EI performance.

Liang Hongsong et al. [22] The paper builds upon the existing literature on organizational innovation motivation and employs OIMS (Organizational Innovation Motivation Scale) questionnaires for the research. By utilizing factor analysis and other empirical methods, the study analyzes the relationship between inner and outer motivations for organizational innovation.

Ghassan Kbar et al. [23] This paper explores the various factors that influence student motivation, categorizing them as intrinsic and/or extrinsic. The research focuses on assessing multifactor-based motivation to differentiate between positive and negative factors. By identifying and understanding these factors, the aim is to control student motivation and enhance student engagement effectively. This can be achieved by minimizing or eliminating the negative factors while encouraging and improving the positive factors.

K. S. Potapov et al. [24] The analysis conducted in this study involved the utilization of contemporary models of motivation and the exploration of a novel motivation system implementation. This newly developed motivation system holds the potential for practical application within modern technical Institutions of Higher Education. It encompasses both commercial and non-commercial components, offering a comprehensive approach to enhance motivation among students and faculty.

Shen Xilin et al. [25] Effective motivation plays a crucial role in promoting staff enthusiasm, initiative, and creativity within an organization. This paper delves into the significance of implementing staff motivation, drawing insights from research, investigations, and rational analysis. The study thoroughly examines the importance of motivating employees, identifies challenges that may arise during its implementation, and explores the principles for effectively executing motivational strategies. Moreover, the paper seeks to propose relevant countermeasures to address any issues that hinder successful motivation initiatives

Lex van Velsen et al. [26] The process of user video tagging holds significant potential in improving the indexing of vast video collections and offering personalized outputs. User-generated tags can effectively enhance the organization and categorization of videos, making it easier to find relevant content. This indexing aspect seems to be of greater importance to users compared to motivations related to socializing or communication.

Dian Anggraini Kusumajati et al. [27] The research seeks to investigate the impact of lecturers' competencies on student achievement motivation. The study adopts a quantitative-based approach and involves 30 respondents from Bina Nusantara University students. The results reveal that lecturers' competencies have a 34.4% influence on students' performance motivation. However, the majority of the contribution, 65.6%, is attributed to other factors beyond lecturers' competencies.

Haiyan et al. [28] The rising trend of private enterprises opting for overseas listings has drawn considerable attention to the study of their motivations behind this decision. This paper presents a set of hypotheses concerning the motivations driving private enterprises to pursue overseas listings. Subsequently, the study employs empirical research methods, including data collection and variable selection, to establish a Logit model.

Bustos et al. [29] The primary objective of this research is to explore and establish the correlation between student motivation in an online course of Computer Programming in Python and various sociodemographic variables and school-related factors, particularly within the context of the pandemic. The study aims to uncover how different factors, such as age, gender, educational background, and school-related variables, may influence students' motivation levels in the specific online course.

Matija Vigato et al. [30] very second, millions of gamers around the world engage in playing video games, driven by a diverse range of reasons. While video games are primarily designed for entertainment, their application extends beyond this realm. Increasingly, they are being utilized for educational, artistic, and therapeutic purposes. Despite these varying purposes, the motivations behind players' engagement with video games are vast and multifaceted.

Lopez et al. [31] Convolutional Neural Networks (CNNs) have historically been closely associated with Computer Vision tasks, playing a pivotal role in significant advancements in Image Classification and serving as the foundational component of many Computer Vision systems. However, in recent times, CNNs have demonstrated their versatility by being applied to problems in Natural Language Processing (NLP) with intriguing outcomes. This paper aims to elucidate the fundamentals of CNNs, exploring their various variations, and delving into their applications in NLP.

Strubell et al. [32] This paper highlights a crucial concern regarding the substantial costs associated with training and developing advanced neural network models for NLP tasks. These models incur significant financial expenses due to the high costs of hardware, electricity, and cloud computing time required for their training. Moreover, the environmental impact of these resource-intensive processes cannot be overlooked, as the carbon footprint from powering modern tensor processing hardware is substantial.

Otter et al. [33] In recent years, the domain of natural language processing has witnessed significant advancements, primarily driven by the widespread adoption of deep learning models. This article offers a concise introduction to the field of natural language processing, providing a rapid overview of deep learning architectures and methodologies. Subsequently, it comprehensively reviews numerous recent studies and compiles a diverse range of significant contributions in the field.

Al-Ayyoub et al. [34] NLP tasks are gaining significant prominence in Online Social Networks (OSN), and Deep Learning (DL) has opened up exciting new avenues for researchers and practitioners to tackle these tasks effectively. This paper presents a comprehensive survey of published works that employ DL techniques for NLP, with a specific focus on the Arabic language. Despite the widespread adoption of social networks in the Arab world, DL techniques have not received the attention they deserve from the Arabic NLP (ANLP) community, in contrast to the attention given to other languages. The paper highlights the existing gap and emphasizes the need for further exploration and utilization of DL in ANLP research.

Wu et al. [35] In the realm of Clinical Natural Language Processing (NLP), the adoption of Deep Learning (DL) techniques has seen a remarkable surge, with publications more than doubling each year until 2018. Among the various DL methods utilized, Recurrent Neural Networks (RNNs) held the top spot at 60.8%, closely followed by word2vec embedding at 74.1%. The predominant focus in information extraction tasks revolved around text classification, named entity recognition, and relation extraction, encompassing an impressive 89.2% of the research landscape.

Young et al. [36] Deep learning methods have revolutionized various domains by leveraging multiple processing layers to learn hierarchical representations of data, leading to state-of-the-art results. In the field of natural language processing (NLP), there has been an explosion of diverse model designs and methodologies. This paper presents a comprehensive review of significant deep learning-related models and methods that have been applied to numerous NLP tasks.

Ramaswamy et al. [37] Accurate analysis of such data is of utmost importance as it unveils valuable insights ranging from consumer buying patterns to product weaknesses, granting a significant competitive edge to businesses. Furthermore, this analysis presents a golden opportunity to unveil customer interests, identify areas for product enhancements, and gain valuable marketing insights. In this paper, we delve into the exploration of various technologies encompassing Deep Learning and Natural Language Processing (NLP). By harnessing these cutting-edge technologies, we aim to enhance the analysis of contextual information and effectively capture customer feedback.

Torfi et al. [38] Natural Language Processing (NLP) plays a crucial role in equipping intelligent machines with the ability to comprehend and interact with human language effectively, facilitating seamless communication between humans and computers. Data-driven approaches have become increasingly prevalent, driven by the remarkable advancements achieved through the application of deep learning methods in various domains, including Computer Vision, Automatic Speech Recognition, and, notably, NLP.

Statement of the Problem

In the 21st century, motivation, motivational quotes, and videos have become indispensable for many individuals in society, particularly the younger generation. However, a question arises regarding the necessity of motivation and whether the depth of motivational videos holds any logical value. There are moments when we all require a dose of motivation, something that can uplift us and reignite our enthusiasm for life. However, it is important to acknowledge that motivational videos may not offer substantial assistance unless we seek momentary upliftment and a temporary boost. In such cases, it is crucial to recognize that our purpose for watching these videos is primarily for entertainment rather than expecting them to bring about life-changing transformations.

When it comes to motivational videos, YouTube undeniably stands out as the premier platform. YouTubers strategically utilize captivating words in their video titles to attract audiences. Within the realm of YouTube, motivational videos can be broadly categorized into two types: intrinsic and extrinsic motivation. These two types encompass distinct factors, behaviors, and theories. Nevertheless, both forms of motivation play a role in our daily lives to varying degrees. Intrinsic motivation refers to the internal drive that compels us to take action without seeking external rewards. On the other hand, extrinsic motivation involves external factors or rewards that motivate us to engage in certain activities. In one way or another, both intrinsic and extrinsic motivation influence our daily lives.

Therefore, we check if there is a solution to the problem that occurs to a YouTube audience from a Motivational Video on YouTube in the perspective of the Intrinsic Motivation video title and Extrinsic Motivation Video Title.

Objective

As mentioned in the earlier problem statement, our objective is to classify YouTube videos into intrinsic and extrinsic motivation categories using their titles and descriptions. In order to achieve this, we aim to evaluate the performance of the dataset by comparing the results obtained from these two types of motivation videos based on their titles and descriptions.

The primary goal of our project is to classify and categorize motivational videos on YouTube based on their titles and descriptions. We aim to analyze and evaluate the performance of these classified YouTube motivational videos to determine the accuracy of our results. By conducting a performance comparison, we seek to assess the effectiveness of our classification approach and provide insights into the accuracy of the categorization process.

So that, the audience or the viewer can make them aware of these two types of videos with the title and description on YouTube Motivational Videos. And also, they can realize the difference between those videos.

Dataset

The initial step involved in this study was the creation of a dataset. However, acquiring video data from YouTube proved to be a challenging task. In order to retrieve the titles and descriptions of the videos, we needed to first obtain the unique IDs of the desired videos [3]. These Video IDs can be obtained from the search results page. To streamline this procedure, we made use of different APIs, including the YouTube Data API v3 provided by Google. Google offers a diverse range of APIs, each designed for specific applications in various fields. By utilizing the YouTube Data API v3, we were able to access functionalities that facilitated the extraction of relevant data, thereby simplifying the data collection process for our project.

- Video searching on YouTube,
- Retrieve information on videos from YouTube either of channel or videos such as likes/dislikes, comments, etc.
- Can start the YouTube video directly from the application.

As an alternative to using APIs for data collection, a manual approach can be adopted to gather the required information and create the dataset. This method entails individually watching YouTube motivational videos and collecting the necessary data. It is important to note that this manual approach requires additional time and effort. While watching the videos, the task involves evaluating and determining whether the video titles reflect intrinsic or extrinsic motivation. Subsequently, the videos must be categorized accordingly into their respective types: videos with titles representing the content and videos with titles indicating extrinsic motivation. By following this process, a dataset can be created that accurately represents the different types of YouTube motivational videos.

Sample Dataset

Id	Video Id	Title	Description	Category
1	96iaZxKRmK	LISTEN TO THIS EVERY NIGHT Before You Sleep | Peaceful Night Affirmations By Sandeep Maheshwari	LISTEN TO THIS EVERY NIGHT Before You Sleep! Sandeep Maheshwari is a name among millions who struggled, failed	Intrinsic
2	iWop_FJaSY	Vivek Oberoi with Dr Vivek Bindra | Devendra Patel CP Bada Business	Devendra Patel Authorized Channel Partner – Bada Business Pvt LTD an Initiative by Dr. Vivek Bindra	Extrinsic
3	RySO0QgoH	HARD WORK PAYS OFF- Best Study Motivation	HARD WORK PAYS OFF! Keep reading. Keep studying. Keep taking care of yourself.	Intrinsic
4	F8e3u7tAx11	HOW I Started My Business | Success Story | Vivek Oberoi |Dr Vivek BIndra	Dr. Vivek Bindra is the founder and CEO of Bada Business Pvt Ltd. One of The Most Progressive Ed-Tech platforms in South East…	Extrinsic
5	8aoXddyBmf	Think Right, Speak Right & Do Right By Gaur Gopal Das	Think Right, Speak Right & Do Right By Gaur Gopal Das.	Intrinsic
6	Uk3HqGX2y	Best Motivational Speech For Hustlers And Future Entrepreneurs | Motivational Madness |	Best Motivational Speech For Hustlers And Future Entrepreneurs	Extrinsic
7	E6NiUExHzO	Life Is Like A Candle 🕯 - Gaur Gopal Das	Gaur Gopal Das is an Indian monk, lifestyle coach, motivational speaker and former HP engineer. He is a member of the ...	Intrinsic
8	7BOi0H59tX	Steve Harvey Leaves the Audience SPEECHLESS | One of the Best Motivational Speeches Ever	Steve Harvey, host of Family Feud and The Steve Harvey Morning show, stand-up comedian, and author, delivers one of the best ...	Extrinsic
9	NT8ZUUp5K	Saying SORRY means you are not always WRONG | Gaur Gopal Das | Motivational Quotes	Saying SORRY means you are not always WRONG | Gaur Gopal Das | Motivational Quotes Motivational Speaker ...	Intrinsic
10	X9JExlvPwcs	Jack Ma Motivational Video | Believe In Your Dreams | Inspirational Speech | Startup Stories	Startup Stories presents, Jack Ma's Motivational Video. In this inspirational speech, Jack Ma, the founder of Alibaba and AliPay ...	Extrinsic

Model Evaluation

In my project, I am leveraging the power of both machine learning and deep learning techniques to tackle complex challenges and achieve impactful results. Machine learning enables me to build predictive models and make data-driven decisions by learning patterns and relationships from labeled datasets. On the other hand, deep learning empowers me to work with unstructured data such as text, by employing neural networks with multiple layers to extract high-level features. By combining the strengths of both approaches, I can address a wide range of problems and explore the potential of artificial intelligence in solving real-world problems efficiently and effectively. The synergy between machine learning and deep learning allows me to develop sophisticated models and gain deeper insights, making my project both versatile and promising in its outcomes.

A. Machine Learning Models

Our research focuses on binary classification, which involves categorizing data into two distinct categories. To achieve this, we will utilize a range of algorithms, including Support Vector Machine (SVM), Logistic Regression, Multinomial Naive Bayes, and Random Forest. These models will be trained using the preprocessed data as input.

Support Vector Machine Algorithm

Support Vector Machine (SVM) is a widely used supervised learning algorithm, commonly applied to both classification and regression tasks. However, it is particularly renowned for its effectiveness in solving classification problems within the field of machine learning.

The primary goal of the SVM algorithm is to find the optimal decision boundary, known as the hyperplane, that can effectively segregate the data points in an n-dimensional space into their respective classes. By establishing this decision boundary, future data points can be accurately classified into the appropriate categories.

SVM achieves this by selecting the most significant points or vectors that assist in constructing the hyperplane. These critical instances are referred to as support vectors, hence lending the algorithm its name, Support Vector Machine.

Logistic Regression

Logistic regression is a widely popular machine learning algorithm categorized under supervised learning. It is primarily employed for predicting a categorical dependent variable based on a given set of independent variables. The main objective of logistic regression is to predict the output of a categorical dependent variable. As a result, the outcome is represented by discrete or categorical values, such as Yes or No, 0 or 1, or True or False.

Unlike providing exact 0 or 1 values, logistic regression generates probabilistic values that range between 0 and 1. It accomplishes this by fitting an "S"-shaped logistic function instead of a regression line, enabling predictions of two maximum values (0 or 1). While linear regression is utilized for solving regression problems, logistic regression is specifically designed for addressing classification tasks.

Logistic regression is a valuable machine learning algorithm as it can provide probabilities and classify new data using both continuous and discrete datasets. It is capable of classifying observations using various types of data, providing insights into the most influential variables used for classification purposes.

Multinomial Naive Bayes

The Multinomial Naive Bayes algorithm is a probabilistic learning method frequently used in Natural Language Processing (NLP) tasks. Based on the Bayes theorem, this algorithm's objective is to predict the tag or category of a given text, such as an email or newspaper article. The algorithm calculates the probability of each tag for a specific sample and selects the tag with the highest probability as the output.

The Naive Bayes classifier comprises multiple algorithms that share a common principle, which is considering each feature to be classified as independent of any other feature. This implies that the presence or absence of one feature does not influence the presence or absence of another feature.

Naive Bayes is a powerful algorithm frequently employed in text data analysis, especially for problems involving multiple classes. To comprehend the functioning of the Naive Bayes theorem, it is essential to have a foundational understanding of the Bayes theorem, as the former is based on the latter.

Bayes's theorem, formulated by Thomas Bayes, calculates the probability of an event occurring based on the prior knowledge of conditions related to that event. It is expressed by the following formula:

$$P\left(\frac{A}{B}\right) = \frac{P(A) \times P(\frac{B}{A})}{P(B)}$$

Eq. - 1

In Bayes's theorem, we calculate the probability of class A when predictor B is already provided.

P(B) = prior probability of B

P(A) = prior probability of class A

P(B/A) = occurrence of predictor B given class A probability

Random Forest Algorithm

Random Forest is a widely popular machine learning algorithm that belongs to the category of supervised learning. Random Forest is capable of handling both classification and regression problems in machine learning. The algorithm is based on the concept of ensemble learning, which entails combining multiple classifiers to address complex problems and improve the overall performance of the model.

As the name implies, Random Forest comprises multiple decision trees built on different subsets of the given dataset. The algorithm aggregates the predictions from each individual tree and determines the final output through a majority voting scheme.

B. Deep Learning Models

Model evaluation in deep learning involves assessing the performance and effectiveness of a trained neural network model. We will use some algorithms such as one is simple Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), BI-LSTM, Convolution Neural Networks (CNN), etc. The pre-processed data will be fed to these models.

Recurrent Neural Network (RNN)

A Recurrent Neural Network (RNN) is a specialized type of artificial neural network designed to process sequential data. It leverages recurrent connections that enable it to handle sequences or time-series data efficiently. RNNs are particularly effective for tasks where the current input relies not only on the present state but also on previous inputs and their corresponding states.

Unlike feedforward neural networks that have a unidirectional flow of information, RNNs have connections that form a cyclic structure, allowing them to retain information about past inputs. This cyclic structure enables RNNs to model temporal dependencies and capture patterns in sequential data.

The basic component of an RNN is the recurrent unit, typically a simple neural network layer with an added recurrent connection. One of the most widely used recurrent units is the Long Short-Term Memory (LSTM) unit. LSTMs address the vanishing gradient problem by incorporating memory cells and gates that control the flow of information.

On the other hand, the kernel trick is a technique employed by Support Vector Machines (SVMs) to implicitly map input features into a higher-dimensional feature space. This mapping facilitates the handling of complex decision boundaries without explicitly calculating the transformed feature vectors. The kernel trick enhances the capabilities of SVMs and enables them to effectively deal with intricate patterns and non-linear relationships in the data.

Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specific type of recurrent neural network (RNN) architecture that addresses the challenge of vanishing gradients and effectively captures long-term dependencies in sequential data. LSTMs are highly suitable for tasks involving the processing and modeling of sequential data, including speech recognition, language modeling, machine translation, and sentiment analysis.

The significant advancement of LSTM networks lies in their capability to selectively retain or forget information across multiple time steps, achieved through specialized memory cells and gating mechanisms. These mechanisms enable LSTMs to retain important information over extended sequences and overcome the issue of vanishing gradients encountered by traditional RNNs. By incorporating memory cells and gates, LSTMs effectively capture and leverage long-term dependencies, resulting in improved performance and enhanced modeling of sequential data.

Main components of an LSTM unit are as follows:

Cell State (CT):
The cell state acts as the long-term memory of the LSTM unit. It can carry information across many time steps, allowing the network to capture dependencies over extended sequences.

Input Gate (i):
The input gate determines how much of the current input should be stored in the cell state. It calculates a "gate" value between 0 and 1 based on the input and previous hidden state, which controls the amount of information to be added to the cell state.

Forget Gate (f):
The forget gate determines how much of the previous cell state should be forgotten or discarded. It computes a "forget" value between 0 and 1 based on the input and previous hidden state, which determines the amount of information to be removed from the cell state.

Output Gate (o):
The output gate determines how much of the cell state should be exposed to the output. It calculates an "output" value between 0 and 1 based on the input and previous hidden state, controlling the amount of information to be passed from the cell state to the output.

Hidden State (h):
The hidden state represents the short-term memory of the LSTM unit. It is calculated based on the current input, the previous hidden state, and the current cell state. The hidden state is used to make predictions or generate outputs.

During the training process, the LSTM parameters (weights and biases) are learned using backpropagation through time, which extends the backpropagation algorithm to handle sequential data. The gradients are computed and used to update the parameters, allowing the LSTM to learn patterns and relationships within the sequential data.

The LSTM architecture has proven to be effective in capturing long-term dependencies, modeling complex sequences, and achieving state-of-the-art results in various tasks involving sequential data.

Bi-Directional Long Short-Term Memory (BI-LSTM)

BI-LSTM, short for Bidirectional Long Short-Term Memory, is an extension of the Long Short-Term Memory (LSTM) architecture that introduces the ability to capture information from both past and future contexts. While traditional LSTMs process sequential data in a unidirectional manner, from past to future, BI-LSTMs incorporate bidirectional connections to simultaneously consider information from both directions.

In a BI-LSTM, the input sequence is processed by two distinct LSTM layers: one layer reads the sequence in the forward direction, and the other layer reads it in the backward direction. Each layer maintains its own hidden state and cell state. The forward layer processes the sequence from the first element to the last, while the backward layer processes it from the last element to the first. This bidirectional processing empowers the model to capture dependencies from both past and future contexts, enhancing its ability to comprehend and represent the input sequence comprehensively.

BI-LSTMs find widespread application in tasks where context from both directions is crucial, including part-of-speech tagging, named entity recognition, sentiment analysis, and machine translation. By incorporating information from both directions, BI-LSTMs excel at capturing dependencies that might be overlooked by unidirectional models, leading to improved performance and more accurate representations of sequential data.

Convolution Neural Network (CNN)

A Convolutional Neural Network (CNN) is a deep learning neural network that excels at processing and analyzing structured grid-like data, such as images and videos. CNNs are particularly well-suited for computer vision tasks, including image classification, object detection, and image segmentation.

The fundamental concept behind CNNs is the utilization of convolution, where a set of learnable filters or kernels is applied to the input data. These filters perform local operations, extracting relevant features by convolving over the input data using a sliding window approach. The result of this convolutional operation is a feature map that captures various aspects or patterns present in the input data. This ability to extract and identify meaningful features is what makes CNNs highly effective in computer vision tasks.

Main components of an LSTM unit are as follows:

Convolution Layers

These layers consist of multiple filters that extract different features from the input data. Each filter scans the input spatially, capturing local patterns and producing a corresponding feature map.

Pooling Layers

Pooling layers reduce the spatial dimensionality of the feature maps by downsampling. Max pooling is a commonly used technique, where the maximum value within a pooling window is selected as the representative value for that region. Pooling helps in reducing computational complexity, providing translation invariance, and extracting the most salient features.

Activation Functions

Activation functions introduce non-linearities to CNN. The Rectified Linear Unit (ReLU) is frequently used, as it efficiently handles the vanishing gradient problem and introduces non-linearity to the network.

Fully Connected Layers

These layers process the extracted features and perform high-level reasoning by connecting all neurons from the previous layer to every neuron in the current layer. Fully connected layers are typically employed in the final stages of a CNN for classification or regression tasks.

Proposed Method

The proposed method in this project is a hybrid approach that combines the power of machine learning and deep learning techniques. By integrating these two methodologies, we aim to achieve superior performance and tackle complex challenges effectively. Machine learning algorithms are utilized to preprocess and engineer the data, extract relevant features, and build predictive models based on labeled datasets. On the other hand, deep learning comes into play for handling unstructured data, such as text. Deep neural networks with multiple layers are employed to automatically learn and extract intricate patterns and representations from the raw data. By leveraging the strengths of both machine learning and deep learning, our approach can handle diverse data types and uncover intricate relationships within the data, ultimately leading to more accurate and robust predictions. This hybrid method provides a comprehensive and versatile solution, promising exciting advancements and breakthroughs in our project's domain.

A. Machine Learning Methodology

In this stage, we leverage Natural Language Processing (NLP) or Text Mining techniques. NLP encompasses several crucial steps, which include the following:

Remove Accented Characters

This step is crucial as it involves converting all characters, including accented characters, into a machine-readable format. Accented characters, such as â, î, or ô, have diacritics above the characters that can pose challenges for processing and analysis. By converting these accented characters into a standardized format, further measures and processing can be implemented more easily.

Case Conversion

The subsequent critical step in the sequence entails converting the text's case. Case sensitivity is a crucial factor in NLP as lowercase and uppercase letters are considered distinct by machines. Therefore, it is necessary to convert the case of the text into either lowercase or uppercase to ensure consistency and accurate analysis.

Reducing Repeated Characters and Removal of Punctuation

This step is crucial because there may be instances where characters are repeated excessively, which cannot be easily detected by a spell-checker at a later stage. Therefore, it is necessary to address this scenario before applying a spell-checker function. Additionally, there may be cases where punctuations are repeated unnecessarily, and it is important to handle them as well.

Remove Special Characters and Numbers

In this step, we will go to remove special characters, as well as we also remove the numerical characters or numbers from the text.

Remove Stopwords

During tokenization, text summarization, text classification, or similar tasks, it is essential to remove stopwords. Doing so allows us to comprehend the context of the textual data more effectively. Removing stopwords is necessary to reduce their influence on the analysis.

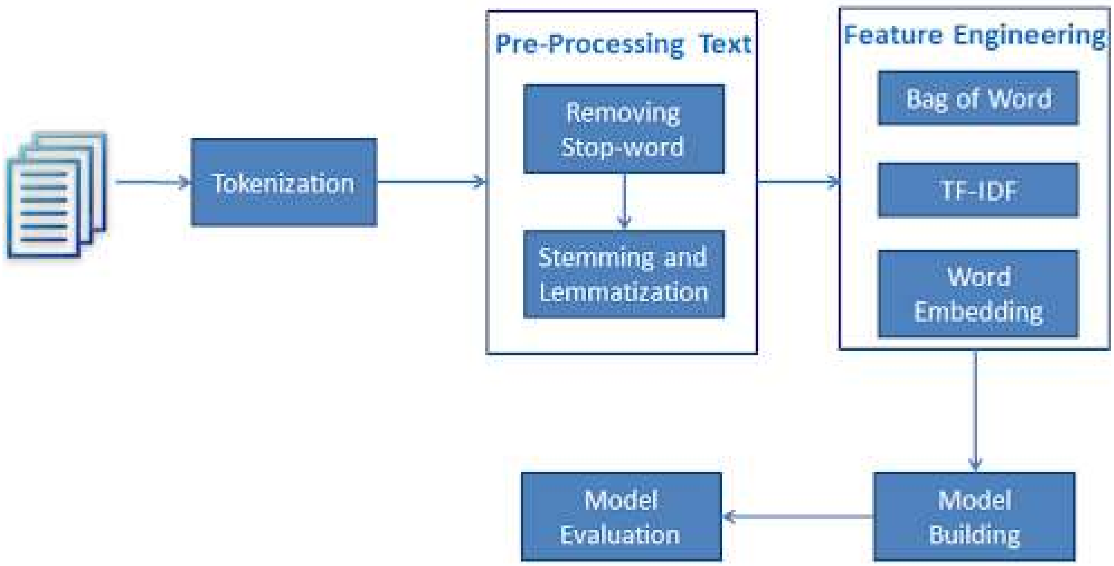
Correcting Mis-spelled words

In this step, we focus on correcting words that have spelling mistakes, which is crucial to ensure accuracy in subsequent steps of the process. It is important to address spelling errors because they can significantly impact the meaning and interpretation of the text. By correcting these mistakes, we improve the overall quality and reliability of the text data.

Lemmatization/Stemming

Lemmatization and stemming both involve reducing words to their root form, such as transforming "planning" to "plan." However, stemming cannot handle certain words like "better," whereas lemmatization can accurately convert it to its root form, "good." This distinction highlights the effectiveness of lemmatization while stemming has been used only in specific instances during our work.

Ultimately, the dataset will undergo preprocessing to generate a clean text. This process will yield a preprocessed dataset that is both labeled and ready for input into the models.



Model Architecture

Figure - 1

B. Deep Learning Methodology

In deep learning for Natural Language Processing (NLP) tasks, there are additional steps specific to language processing. Here is a summary of the procedure employed in deep learning for NLP:

Data Collection

Collect a substantial dataset that comprises relevant text data pertaining to the specific NLP task. This dataset may consist of various forms of textual information, such as text documents or individual sentences.

Text Preprocessing

To prepare the text data for deep learning purposes, data cleaning and preprocessing are performed. This stage encompasses several tasks, including removing punctuation, converting text to lowercase, tokenizing (segmenting) the text into individual words or sub-word units, eliminating stop words, and handling special characters or symbols. These steps are essential to ensure the text data is well-suited for effective deep-learning analysis.

Word Embeddings

Convert words into numerical vectors known as word embeddings, which capture the semantic relationships and contextual information of the words. Popular methods for generating word embeddings include word2vec, GloVe, and the utilization of contextualized word embeddings such as BERT or GPT.

Sequence Padding

In order to feed variable-length sequences into a neural network, it is important to adjust the sequences to a fixed length. This can be achieved through padding or truncation. Padding entails adding zeros or a designated token to shorter sequences, while truncation involves removing excess elements from longer sequences.

Model Architecture Design

Choose a suitable deep learning architecture tailored to the specific NLP task at hand. Options include recurrent neural networks (RNNs) like Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRUs), convolutional neural networks (CNNs), or transformer-based models such as BERT or GPT. The selection of the appropriate architecture is crucial for achieving optimal performance in the NLP task.

Deployment and Inference

Put the trained model into deployment to make predictions on new and previously unseen text data. This process involves passing the new text through the trained network and obtaining predictions based on the patterns learned during training.

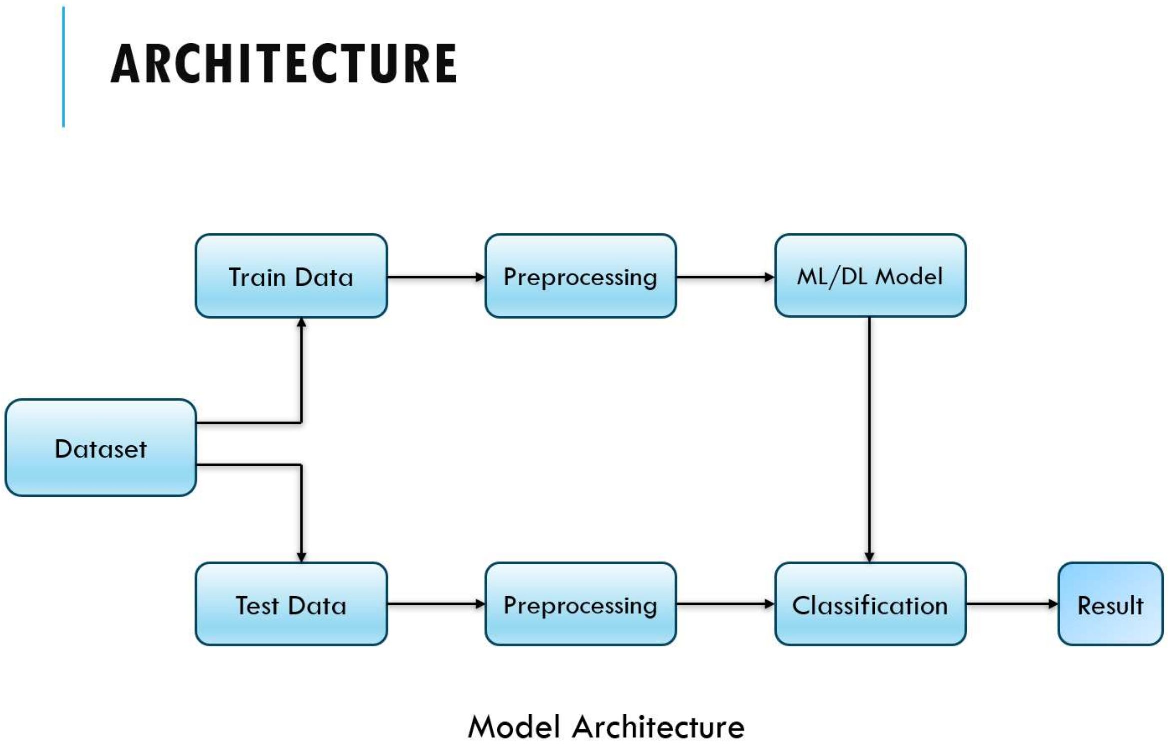
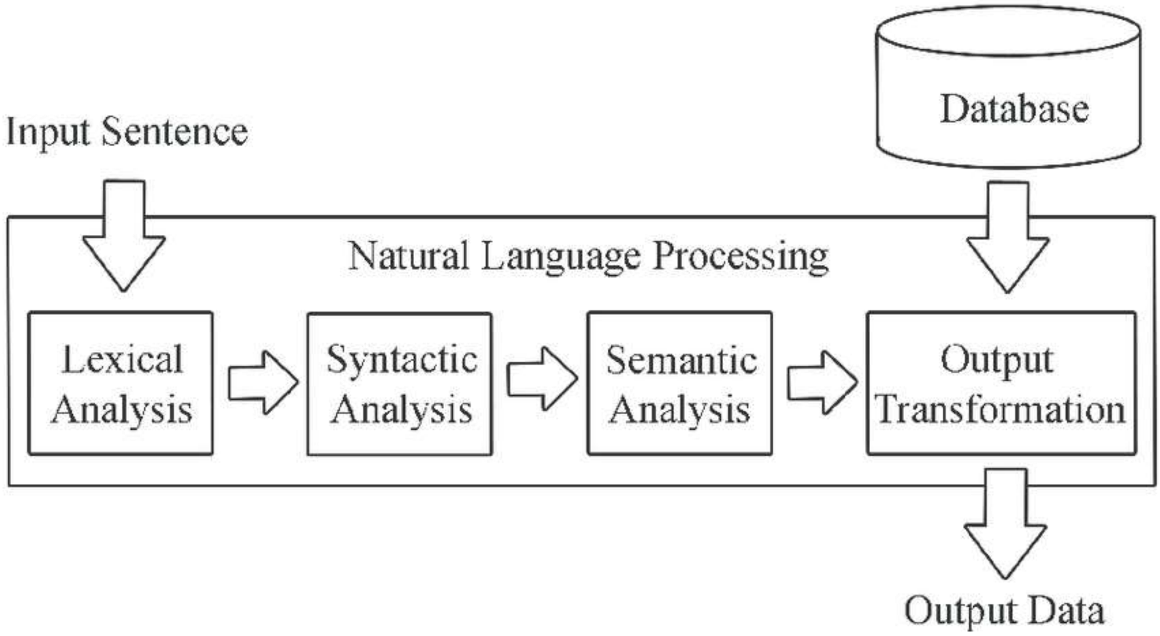


Figure -2

Data Preprocessing

Data Preprocessing for Machine Learning

Data processing involves transforming raw data into usable information for models. In order to make the data suitable for classifiers, we employed data preprocessing techniques. Raw data often contain noise that can negatively impact model accuracy. Through the implementation of data preprocessing, we successfully generated clean text that could be efficiently utilized in our system. Given that the raw data collected from YouTube can be noisy, we applied preprocessing techniques to handle this issue effectively. This involved three stages: tokenization, count vectorization, and TF-IDF. The Keras tokenizer library was employed for tokenization, breaking down video titles into individual tokens. We then employed count vectorization and TF-IDF techniques for further preprocessing. Furthermore, we performed text splitting after tokenization and made use of a label encoder to preprocess the data, preparing it for our proposed model.



Data Preprocessing using NLP

Figure - 3

TF-IDF

TF-IDF, which stands for Term Frequency - Inverse Document Frequency, is a widely used statistical technique in natural language processing and information retrieval. TF-IDF evaluates the importance of a term in a document compared to a collection of documents, which is commonly referred to as a corpus. Text vectorization involves transforming words in a text document into numerical values, effectively representing their significance or importance. TF-IDF is one of the commonly used scoring schemes in text vectorization. TF-IDF earns its name by calculating a score for each word by multiplying its Term Frequency (TF) with the Inverse Document Frequency (IDF).

Term Frequency

The Term Frequency (TF) of a word or term is computed by dividing the number of times the word appears in a document by the total word count in that document.

$$TF = \frac{Numbers\ of\ time\ the\ term\ appears\ in\ the\ document}{Total\ number\ of\ terms\ in\ the\ document}$$

Eq. - 2

Inverse Document Frequency

The Inverse Document Frequency (IDF) of a term signifies the proportion of documents in the corpus that contain that specific term. Words that are unique to a small percentage of documents, like technical jargon terms, receive higher IDF values, indicating their significance, whereas words that are common across all documents, such as "a," "the," and "and," receive lower IDF values.

$$IDF = \log\left(\frac{Number\ of\ the\ documents\ in\ the\ corpus}{Number\ of\ documents\ in\ the\ corpus\ contain\ the\ term}\right)$$

Eq. – 3

The TF-IDF of a term is calculated by multiplying TF and IDF scores.

$$TF\text{-}IDF = TF * IDF$$

Eq. - 4

In simpler terms, a term is considered important if it appears frequently within a specific document but rarely in other documents. In other words, the TF component captures the term's prevalence within the document, while the IDF component measures its uniqueness across the corpus. The resulting TF-IDF score serves as a measure of the term's importance for a particular document within the entire corpus.

TF-IDF is a valuable technique widely used in various natural language processing applications. For example, search engines leverage TF-IDF to assess the relevance and ranking of documents in response to a given query. Moreover, TF-IDF finds application in tasks such as text classification, text summarization, and topic modeling.

It is important to highlight that there exist various approaches for calculating the IDF score. The most common approach involves using either the base 10 logarithm or the natural logarithm. Furthermore, to prevent division by zero, a value of one can be added to the denominator in the calculation. This adjustment ensures that the IDF score can be accurately computed for all terms in the corpus.

$$IDF = \log\left(\frac{Number\ of\ the\ documents\ in\ the\ corpus}{Number\ of\ documents\ in\ the\ corpus\ contain\ the\ term + 1}\right)$$

Eq. - 5

Data Preprocessing for Deep Learning

Data preprocessing holds a crucial role in deep learning, as it transforms raw data into a suitable format that can be effectively utilized by a deep learning model. This process encompasses several preprocessing steps, including data cleaning, normalization, feature scaling, encoding categorical variables, and dividing the data into training, validation, and testing sets. These steps are essential for preparing the data in a way that maximizes the model's performance and facilitates accurate and efficient learning.

Data Cleaning

During this step, the focus is on addressing missing values, outliers, and noise present in the dataset. Missing values can be dealt with by either imputing them with appropriate values or removing the affected instances. Outliers can be managed through techniques like trimming or winsorization. Additionally, noise in the data can be reduced through smoothing techniques

Data cleaning is a crucial step in the data preprocessing pipeline, where the focus is on addressing missing values, outliers, and noise within the dataset. This process is essential to guarantee that the data is of high quality, devoid of inconsistencies, and well-suited for subsequent analysis or modeling tasks.

Normalization

Normalizing the data is a critical step to ensure that all input features are brought to a comparable scale, which can lead to improved convergence and performance during model training. Common normalization techniques include min-max scaling, where values are scaled to a predefined range such as [0, 1], and z-score normalization, which scales values to have a mean of zero and a standard deviation of one.

Feature Scaling

Scaling numerical features to a consistent range is beneficial in preventing certain features from overpowering others during the training process. This can be achieved through techniques like standardization (z-score normalization) or min-max scaling.

Encoding Categorical Variables

Categorical variables typically require encoding into numerical values before they can be utilized in a deep-learning model. One-hot encoding and label encoding are widely employed techniques for this purpose.

Splitting the Dataset

A common practice involves splitting the dataset into three distinct subsets: training, validation, and testing. The training set is utilized to train the model, allowing it to learn patterns and relationships within the data. The validation set is used for fine-tuning hyperparameters, monitoring the model's performance, and making necessary adjustments if required.

Word Embedding

Word embedding is a potent technique applied in natural language processing (NLP) and deep learning to represent words as dense vector representations in a high-dimensional space. The primary objective of word embedding is to capture the semantic and syntactic relationships between words, facilitating machines to gain a deeper understanding and process human language more

effectively.

Word embedding techniques, including Word2Vec, GloVe (Global Vectors for Word Representation), and FastText, overcome these limitations by mapping words to dense, continuous vector representations. These vectors are generated by training on extensive collections of text data using unsupervised learning algorithms. Through this process, word embeddings are formed, which position similar words in close proximity within the embedding space.

Word embeddings provide multiple benefits. Firstly, they capture semantic similarity, enabling the identification of related words and concepts. For instance, by performing vector arithmetic operations such as subtracting the vector for "king" from "man" and adding "woman," we can approximate a vector that is close to "queen." Secondly, word embeddings can handle out-of-vocabulary (OOV) words by inferring their representations based on contextual information. This allows the model to represent and understand words not seen during training. Lastly, the continuous vector representations of word embeddings enable machine learning models to process textual data as real-valued input, which leads to efficient computation and training.

GloVe

GloVe (Global Vectors for Word Representation) is a word embedding technique created by researchers at Stanford University. The main goal of GloVe is to produce word embeddings that encapsulate the global statistical properties of words. It achieves this by utilizing co-occurrence statistics derived from a vast corpus of text data.

Unlike certain other word embedding techniques, GloVe effectively combines the strengths of global matrix factorization methods and local context window-based methods. It incorporates both the broader word co-occurrence statistics across the entire corpus and the specific context information within local windows.

The fundamental concept underlying GloVe involves creating a co-occurrence matrix that captures the frequency of word co-occurrences within a defined context window. This matrix is subsequently factorized using matrix factorization techniques like Singular Value Decomposition (SVD) to derive word embeddings.

GloVe embeddings have found widespread application in various natural language processing tasks, including sentiment analysis, text classification, machine translation, and question-answering. The availability of pre-trained GloVe embeddings in different dimensions, such as 50, 100, 200, and 300, provides a valuable resource for researchers and practitioners in the field of NLP.

Evaluation Parameters

Cross-Validation

Cross-validation is a valuable technique employed in deep learning to evaluate and assess the performance and generalization capabilities of predictive models. It involves dividing the available labeled dataset into multiple subsets or folds, with each fold serving as both a training set and a validation set.

The original dataset is divided into k equal-sized folds, typically choosing a value between 5 and 10 for k. Each fold contains an approximately equal distribution of samples from each class, ensuring a representative distribution across the folds.

The cross-validation process is then performed iteratively for k times. In each iteration, one of the k folds is designated as the validation set, while the remaining k-1 folds are used as the training set. The model is trained and evaluated using these sets, providing insights into its performance and generalization across different subsets of the data.

Cross-validation is widely employed in deep learning for robust model evaluation and selection. Assessing performance across multiple folds, it offers a more comprehensive understanding of a model's capabilities, aiding in making informed decisions about model deployment and optimization.

Confusion Matrix

A confusion matrix is a tabular representation that provides a concise summary of a classification model's performance. It presents the count of true positives, true negatives, false positives, and false negatives. The confusion matrix serves as the basis for calculating different evaluation metrics, including accuracy, precision, recall, and F1 score.

Accuracy

Accuracy is a widely recognized and commonly used validation method for evaluating machine learning models, particularly for classification problems. Its popularity stems from its simplicity, as it is straightforward to understand and implement. Accuracy serves as a useful metric to assess model performance, especially in straightforward cases.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Number\ of\ total\ predictions}$$

Eq. - 6

In classification problems, accuracy is a commonly used evaluation metric that indicates the percentage of correct predictions made by a model. It measures the number of correct predictions relative to the total number of predictions made. The accuracy score is computed by dividing the number of correct predictions by the total number of predictions. This metric provides a straightforward measure of how well the model performs in terms of correctly classifying the data.

Accuracy in Binary Classification

We can express accuracy in True/False Positive/Negative values in the binary classification case. The accuracy formula in machine learning is given as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Eq. - 7

Where there are only 2 classes - positive & negative:

TP: True Positives i.e. positive classes that are correctly predicted as positive**.**
FP: False Positives i.e negative classes that are falsely predicted as positive.
TN: True Negatives i.e., negative classes that are correctly predicted as negative.
FN: False Negatives i.e positive classes that are falsely predicted as negative.

Precision

Precision is the first part of the F1 Score. It can also be used as an individual machine learning metric. Its formula is shown here:

$$Precision = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + Number\ of\ False\ Positives}$$

Eq. - 8

Recall

Recall is the second component of the F1 Score, although recall can also be used as an individual machine learning metric. The formula for the recall is shown here:

$$Recall = \frac{Number\ of\ True\ Positives}{Number\ of\ True\ Positives + Number\ of\ False\ Negatives}$$

Eq. - 9

F1 Score

In the realm of classification models, the F1 score serves as a machine learning metric. While several metrics are available for evaluating classification models, this article aims to provide insights into the calculation of the F1 score and the situations in which it provides additional value.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Eq. - 10

Result & Analysis

Result Analysis for Machine Learning

Among the classifiers used, the Support Vector Machine (SVM) achieved the highest accuracy score of 87.89%, while the Random Forest (RF) classifier achieved the lowest accuracy of 80.71%. Additionally, the MNB demonstrated the highest positive predictive value, also known as precision, with a score of 89.65%. The SVM classifier also attained the maximum F1 score value of 79.16%. It is noteworthy that the SVM reached a maximum accuracy of 87.89%.

Result table

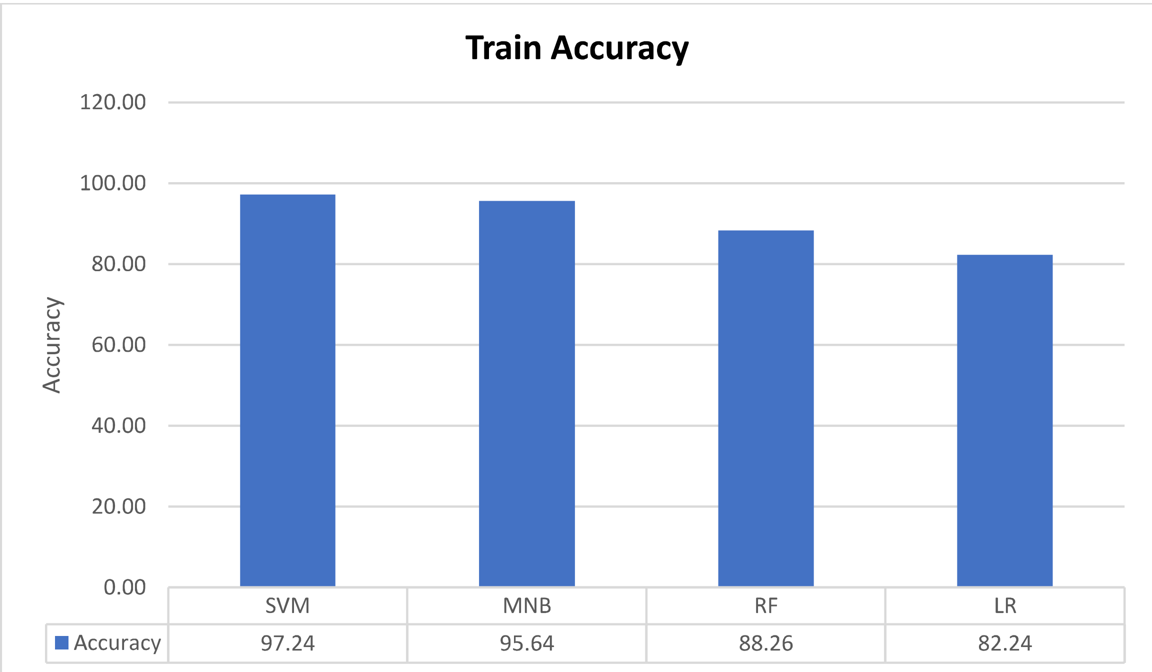
No.	Name	Train Accuracy	Test Accuracy	Accuracy	Precision	Recall	F1 Score
1	**SVM**	97.07%	87.89%	**87.89%**	87.32%	84.32%	79.16%
2	**MNB**	95.64%	84.11%	**84.11%**	89.65%	85.61%	72.95%
3	**LR**	88.26%	83.61%	**83.61%**	80.13%	74.63%	74.04%
4	**RF**	82.24%	80.71%	**80.71%**	77.10%	70.01%	73.32%

Training Accuracy

In the above table, we have the training accuracy of four different machine learning models: SVM (Support Vector Machine), MNB (Multinomial Naive Bayes), LR (Logistic Regression), and RF (Random Forest). Training accuracy represents the percentage of correctly classified samples from the training data during the learning process. A higher training accuracy indicates that the model has successfully learned and recognized patterns and features present in the training set.

SVM achieved an impressive training accuracy of 97.07%, indicating that around 97.07% of the samples in the training data were correctly classified by the SVM model. MNB obtained a slightly lower but still substantial training accuracy of 95.64%, signifying that approximately 95.64% of the samples in the training data were accurately classified by the MNB model. LR achieved an 88.26% training accuracy, showing that about 88.26% of the samples in the training data were correctly classified by the LR model. Lastly, RF achieved a training accuracy of 82.24%, indicating that around 82.24% of the samples in the training data were correctly classified by the RF model.

While higher training accuracy generally suggests that the models have learned well from the training data, it is important to assess their performance on a separate test dataset to ensure their ability to generalize to new, unseen data and avoid overfitting. An in-depth evaluation of test accuracy and other metrics is necessary to determine the overall effectiveness and reliability of each model in practical applications.

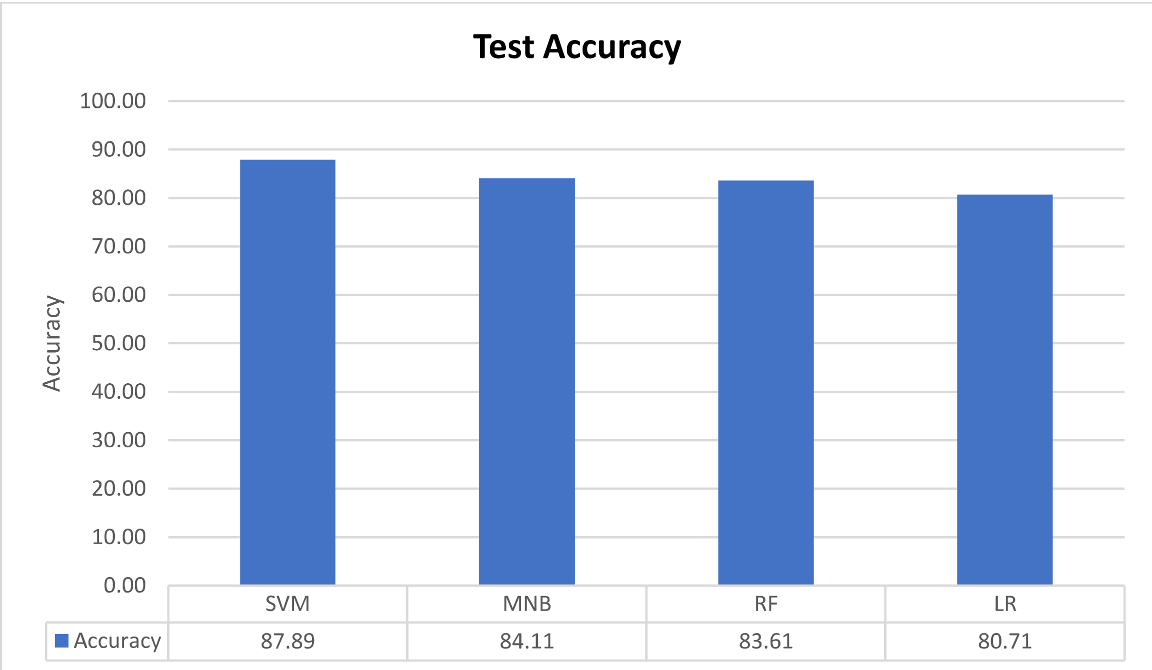


Training Accuracy in Machine Learning

Figure - 4

Testing Accuracy

Testing accuracy in machine learning refers to the percentage of correctly classified samples from a separate test dataset that the model has not seen during the training process. It is a critical metric for assessing the model's generalization ability. Looking at the table, we observe the testing accuracy of four machine learning models: SVM, MNB, LR, and RF. SVM achieved a testing accuracy of 87.89%, indicating that approximately 87.89% of the samples in the test data were correctly classified by the SVM model. MNB obtained a testing accuracy of 84.11%, signifying that around 84.11% of the test samples were accurately classified by the MNB model. LR achieved an 83.61% testing accuracy, while RF achieved 80.71%. The testing accuracy provides valuable insights into how well these models perform on unseen data, allowing us to make informed decisions about their suitability for real-world applications.



Testing Accuracy in Machine Learning

Figure - 5

Now, we evaluate the ROC (Receiver Operating Characteristics) curve to compare the True Positive Rate and False Positive Rate to investigate the used Machine Learning Model.

An **ROC curve** (**receiver operating characteristic curve**) is a graph showing the performance of a classification model at all classification thresholds. This curve plots two parameters:

True Positive Rate (**TPR**) is a synonym for recall and is therefore defined as follows:

$$TPR = \frac{TP}{TP + FN}$$

Eq. - 11

False Positive Rate (**FPR**) is defined as follows:

$$FPR = \frac{FP}{FP + TN}$$

Eq. - 12

ROC Curve

Figure – 6

Result Analysis for Deep Learning

Among the classifiers used, the Convolutional Neural Network (CNN) achieved the highest accuracy score of 88.07%, while the Simple Recurrent Neural Network (RNN) achieved the lowest accuracy of 81.94%. Additionally, the Support Vector Machine (SVM) demonstrated the highest positive predictive value, also known as precision, with a score of 72.69%. The Long Short-Term Memory (LSTM) classifier attained the maximum F1 score value of 66.36%. Notably, the SVM also reached a maximum accuracy of 88.07%.

Result table:

No.	Name	Train Accuracy	Test Accuracy	Accuracy	Precision	Recall	F1 Score
1	**RNN**	97.19%	81.70%	**81.94%**	64.94%	71.09%	65.99%
2	**LSTM**	95.38%	86.38%	**86.51%**	70.02%	68.31%	66.36%
3	**BI-LSTM**	97.45%	84.87%	**85.25%**	70.07%	61.39%	60.64%
4	**CNN**	97.36%	88.67%	**88.07%**	72.69%	65.69%	63.82%

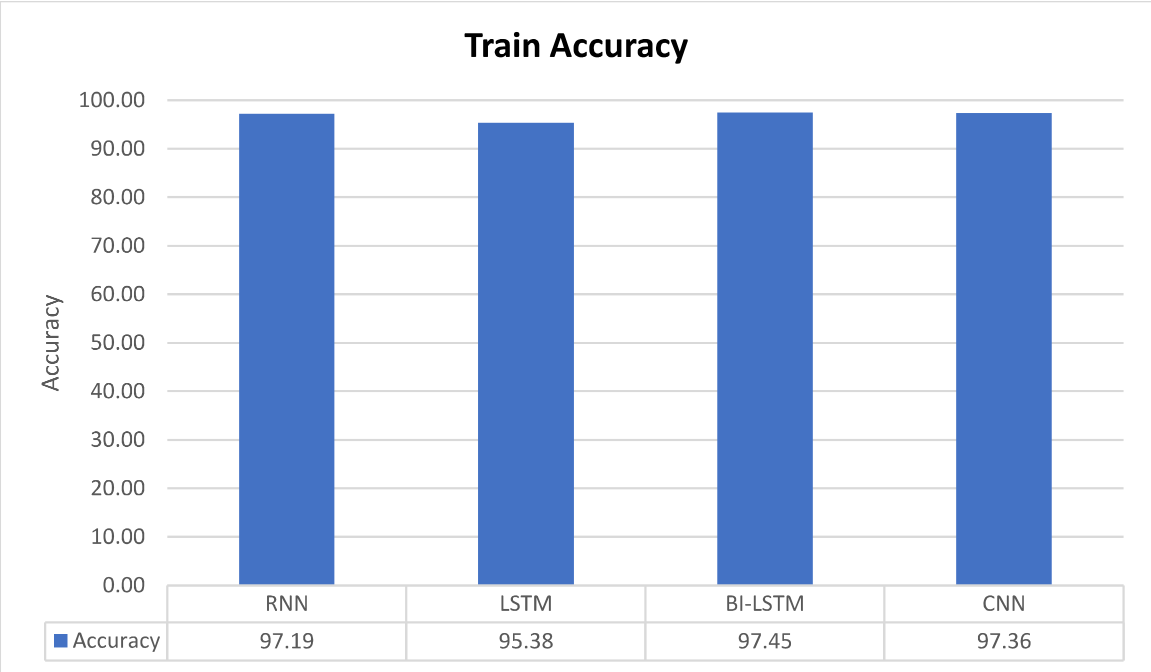
Train Accuracy

The table displays the training accuracy of four different neural network models: RNN (Recurrent Neural Network), LSTM (Long Short-Term Memory), BI-LSTM (Bidirectional LSTM), and CNN (Convolutional Neural Network). Training accuracy represents the percentage of correctly classified samples from the training data during the training process of each model.

RNN achieved a high training accuracy of 97.19%, indicating that approximately 97.19% of the samples in the training data were correctly classified by the RNN model. LSTM obtained a slightly lower but still commendable training accuracy of 95.38%, signifying that around 95.38% of the samples in the training data were accurately classified by the LSTM model. BI-LSTM achieved a remarkable training accuracy of 97.45%, showing that about 97.45% of the samples in the training data were correctly classified by the BI-LSTM model. Finally, CNN achieved an impressive training accuracy of

97.36%, indicating that around 97.36% of the samples in the training data were correctly classified by the CNN model.

The high training accuracy of these models suggests that they have effectively learned from the training data and captured the underlying patterns and features. However, it is essential to evaluate their performance on unseen test data to determine their ability to generalize to new data and assess their real-world effectiveness.



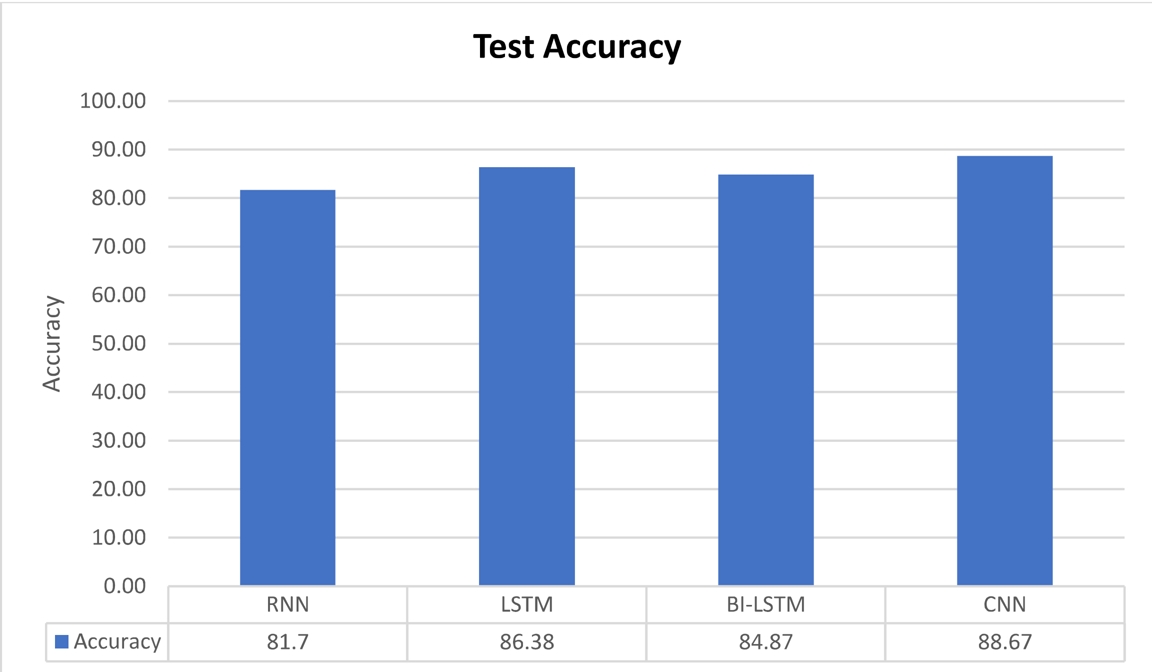
Training Accuracy in Deep Learning

Figure – 7

Testing Accuracy

Here in the deep learning result table, RNN achieved a testing accuracy of 81.70%, indicating that approximately 81.70% of the samples in the test data were correctly classified by the RNN model. LSTM obtained a higher testing accuracy of 86.38%, signifying that around 86.38% of the test samples were accurately classified by the LSTM model. BI-LSTM achieved a testing accuracy of 84.87%, demonstrating that approximately 84.87% of the test samples were correctly classified by the BI-LSTM model. Lastly, CNN achieved an impressive testing accuracy of 88.67%, indicating that around 88.67% of the test samples were correctly classified by the CNN model.

The testing accuracy is crucial as it assesses the models' generalization ability to perform well on new, unseen data. The higher the testing accuracy, the more reliable the model's performance in real-world applications. It is essential to compare the testing accuracy with the training accuracy to ensure that the models have not overfit to the training data. A good balance between training and testing accuracy indicates that the models have learned meaningful patterns and can generalize effectively to new data.



Testing Accuracy in Deep Learning

Figure – 8

Conclusion

In conclusion, the project on analyzing motivational videos using YouTube descriptions has been a significant endeavor in understanding the landscape of inspirational content on the platform. By leveraging machine learning and natural language processing techniques, and deep learning we have been able to extract valuable insights from the vast amount of textual information present in video descriptions. This analysis has shed light on the prevalence of motivational content on YouTube and the diverse themes and topics covered by creators in this genre. Through sentiment analysis and topic modeling, we have identified the emotional impact and key subject areas that resonate with audiences. This project has not only provided valuable knowledge about motivational videos but has also demonstrated the potential of data-driven approaches in understanding user preferences and content trends on one of the world's largest video-sharing platforms. As YouTube continues to evolve and shape digital content consumption, the findings from this analysis contribute to a deeper understanding of the role of motivational videos in today's digital landscape and offer valuable insights for content creators and viewers alike.

References

[1] Alexa. The top 500 sites on the web Global. Retrieved from https://www.alexa.com/topsites. 2019.

[2] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based Natural Language Processing [review article]. IEEE Computational Intelligence Magazine, 13(3), 55-75. https://doi.org/10.1109/mci.2018.2840738

[3] Islam, M., Ria, N. J., Mohammad Masum, A. K., & Ani, J. F. (2021). Performance comparison of multiple supervised learning algorithms for YouTube exaggerated Bangla titles classification. *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. https://doi.org/10.1109/icccnt51525.2021.9579582

[4] Kalra, G. S., Kathuria, R. S., & Kumar, A. (2019). YouTube video classification based on title and description text. *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. https://doi.org/10.1109/icccis48478.2019.8974514

[5] Ni, B., Song, Y., & Zhao, M. (2011). YouTubeEvent: On large-scale video event classification. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. https://doi.org/10.1109/iccvw.2011.6130430

[6] Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based Natural Language Processing [review article]. IEEE Computational Intelligence Magazine, 13(3), 55-75. https://doi.org/10.1109/mci.2018.2840738

[7] Cunha, A. A., Costa, M. C., & Pacheco, M. A. (2019). Sentiment analysis of YouTube video comments using Deep Neutral Networks. Artificial Intelligence and Soft Computing, 561-570. https://doi.org/10.1007/978-3-030-20912-4_51

[8] M. Z. Asghar, S. Ahmad, A. Marwat, and F. M. Kundi, "Sentiment Analysis on YouTube: A Brief Survey," arXiv.org, 30-Nov-2015.

[9] O. Uryupina, B. Plank, A. Severyn, A. Rotondi, and A. Moschitti, "SenTube: A Corpus for Sentiment Analysis on YouTube Social Media," Proceedings of the Ninth International Conference on Language Resources and Evaluation ({LREC},2014.

[10] Bhuiyan, H., Ara, J., Bardhan, R., & Islam, M. R. (2017). Retrieving YouTube video by sentiment analysis on User Comment. *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*. https://doi.org/10.1109/icsipa.2017.8120658

[11] Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment analysis of YouTube movie trailer comments using Naïve Bayes. Bulletin of Computer Science and Electrical Engineering, 1(1), 26-32. https://doi.org/10.25008/bcsee.v1i1.5

[12] Krishna, A. (n.d.). Polarity Trend Analysis of public sentiment on YouTube. https://doi.org/10.31274/etd-180810-146

[13] Porreca, A., Scozzari, F., & Di Nicola, M. (2020). Using text mining and sentiment analysis to analyse YouTube Italian videos concerning vaccination. *BMC Public Health*, *20*(1). https://doi.org/10.1186/s12889-020-8342-4

[14] Singh, S., & Sikka, G. (2021). YouTube sentiment analysis on US Elections 2020. 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC). https://doi.org/10.1109/icsccc51823.2021.9478128

[15] Lavanya, P. M., & Sasikala, E. (2021). Deep learning techniques on text classification using Natural Language Processing (NLP) in Social Healthcare Network: A comprehensive survey. 2021 3rd International Conference on Signal Processing and Communication (ICPSC). https://doi.org/10.1109/icspc51351.9451752

[16] Sharmin, S., & Chakma, D. (2020). Attention-based convolutional neural network for Bangla sentiment analysis. *AI & SOCIETY*, *36*(1), 381–396. https://doi.org/10.1007/s00146-020-01011-0

[17] Aggarwal, N., Agrawal, S., & Sureka, A. (2014). Mining YouTube metadata for detecting privacy invading harassment and misdemeanor videos. 2014 Twelfth Annual International Conference on Privacy, Security and Trust. https://doi.org/10.1109/pst.2014.6890927

[18] Song, J., Lee, S., Kim, B., Seol, S., Lee, B., & Kim, M. (2020). VTIM: Video title identification using open metadata. *IEEE Access*, *8*, 113567–113584. https://doi.org/10.1109/access.2020.3003378

[19] Olivier L. Georgeon and James B. Marshall and Simon Gay. (2012). Interactional Motivation in artificial systems: Between extrinsic and intrinsic motivation. 2012 International Conference on Development and Learning and Epigenetic Robotics https://doi.org/10.1109/devlrn.2012.6400833

[20] Konstantin A. Grebenyuk. (2021). Motivation Generator: An Empirical Model of Intrinsic Motivation for Learning. 2021 International Conference on Engineering, Technology https://doi.org/10.1109/tale52509.2021.9678581

[21] Ying Dong and Qiang Cai. (2012). The impact of motivation on Chinese enterprise. International Conference on Management of Innovation https://doi.org/10.1109/icmit.2012.6225904

[22] Liang Hongsong. (2010), AICI '10: Proceedings of the 2010 International Conference on Artificial Intelligence and Computational Intelligence - Volume 03, October 2010 https://doi.org/10.1109/AICI.2010.332

[23] Ghassan Kbar, Ammar Alazab, Johnson Agbinya. (2019) Multi-factor-based enhancing students' motivations. 2019 International Conference on Industrial Technology https://doi.org/10.1109/icit.2019.8754982

[24] K. S. Potapov, P. V. Terentieva, V. P. Semenov. (2016) About student's scientific research motivation in technical institutions of higher education in Russia. 2016 Russia Young Researchers in Electrical and Electronic Engineering Conference. https://doi.org/10.1109/eiconrusnw.2016.7448311

[25] Shen Xiling, Li Fei (2010) Research on Effective Motivation in Enterprises Administration. International Conference on Information Management, Innovation Management and Industrial Engineering. https://doi.org/10.1109/iciii.2010.254

[26] Lex van Velsen, Mark Melenhorst (2009) Incorporating user motivations to design for video tagging. https://doi.org/10.1016/intcom.2009.05.002

[27] Dian Anggraini Kusumajati, Yustinus Suhardi Ruman, Kristianus Oktriono (2017). The Influence of Lecturers' Competencies Towards Students' Performance Motivation: A Case Study at Higher Education. 2017 International Symposium on Educational Technology. https://doi.org/10.1109/iset.2017.47

[28] Wang Haiyan. (2011). The motivation of private enterprise overseas listing. 2011 International Conference on Consumer Electronics, 2011 Communications and Networks. https://doi.org/10.1109/cecnet.2011.5768800

[29] Solange Barros Bustos, Luis Álvarez-González (2021). Motivation in first year engineering students of the online course "Computer Programming in Python" in the context of a pandemic by COVID-19. 2021 Latin American Conference on Learning Technologies. https://doi..org/10.1109/laclo54177.2021.00017

[30] Matija Vigato, Tihana Babić (2021). Research on Gamer Motivation Factors Based on the Gamer Motivation Model Framework. 2021 44th International Convention on Information, Communication and Electronic Technology. https://doi.org/10.23919/mipro52101.2021.9596942

[31] Lopez, Marc Moreno and Kalita, Jugal (2017) Deep Learning applied to NLP
https://doi.org/10.48550/arXiv.1703.03091

[32] Strubell, Emma and Ganesh, Ananya and McCallum, Andrew (2019) Energy and Policy Considerations for Deep Learning in NLP
https://doi.org/10.48550/arXiv.1906.02243

[33] Daniel W. Otter and Julian R. Medina and Jugal K. Kalita (2021) A Survey of the Usages of Deep Learning for Natural Language Processing
https://doi.org/10.1109/tnnls.2020.2979670

[34] Mahmoud Al-Ayyoub and Aya Nuseir and Kholoud Alsmearat and Yaser Jararweh and Brij Gupta (2018) Deep learning for Arabic (NLP): A survey
https://doi.org/10.1016/j.jocs.2017.11.011

[35] Stephen Wu and Kirk Roberts and Surabhi Datta and Jingcheng Du and Zongcheng Ji and Sarvesh Soni and Qiong Wang and and Yang Xiang and Bo Zhao and Hua Xu (2019) Deep learning in clinical natural language processing: a methodical review. Journal of the American Medical Informatics Association
https://doi.org/10.1093/jamia/ocz200

[36] Tom Young and Devamanyu Hazarika and Soujanya Poria and Erik (2018) Recent Trends in Deep Learning Based Natural Language Processing, IEEE Computational Intelligence Magazine.
https://doi.org/10.1109/mci.2018.2840738

[37] Sridhar Ramaswamy and Natalie DeClerck (2018), Customer Perception Analysis Using Deep Learning and (NLP)
https://doi.org/10.1016/j.procs.2018.10.326

[38] Christopher Manning (2016), Understanding Human Language, Proceedings of the 39th International (ACM) (SIGIR) conference on Research and Development in Information Retrieval
https://doi.org/10.1145/2911451.2926732