A Novel Multimodal Online News Popularity Prediction Model based on Ensemble Learning

Anuja Arora, *Senior Member, IEEE*, Vikas Hassija, Shivam Bansal, Siddharth Yadav, Vinay Chamola *Senior Member, IEEE* and Amir Hussain *Senior Member, IEEE*

***Abstract*—The prediction of news popularity is having substantial importance for the digital advertisement community in terms of selecting and engaging users. Traditional approaches are based on empirical data collected through surveys and applied statistical measures to prove a hypothesis. However, predicting news popularity based on statistical measures applied to past data is highly questionable. Therefore, in this paper, we predict news popularity using machine learning classification models and deep residual neural network models. Articles are usually made up of textual content and in many cases, images are also used. Although it is evident that the appropriate amount of textual data is required to extract features and create models, image data is also helpful in gaining useful information. In this paper, we present a novel multimodal online news popularity prediction model based on ensemble learning. This research work acts as a guide for extensive feature engineering, feature extraction, feature selection, and effective modeling to create a robust news popularity Prediction Model. Three kinds of features – meta-features, text features, and image features are used to design an influential and robust model. The relative error performance measure Root Mean Squared logarithmic error (RMSLE) is used to quantify the popularity prediction error. Further, the RMSLE outcome shows 0.351 which is the lowest error value given by the proposed model. Further, the most important features are also sought out to show the dependence of the best-fit model on text and image features.**

***Index Terms*—ResNet, Deep Residual Neural Network, Ensemble, Feature Engineering, Neural Network, Online news popularity**

I. Introduction

There is a perpetual need and scope for improvement in the advertisement industry especially when it comes to analytics that caters to text content on the web mostly in articles and blogs [1]. These text analytical insights are very valuable for brands and are helpful for digital marketing firms such as news agencies [2], [3], publishers, and brands to boost the popularity of their content and achieve maximum audience engagement. Undoubtedly, for online publishers, news agencies, and brands it is important to produce engaging content among the audience. Audience engagement is always a moving target, and there is more competition that requires every minute of attention. Consistently engaging an audience based on intuition and guesswork is not viable. Additionally, content writers and content managers are unaware about

- What type of content garners maximum traffic?
- Which keywords are loved by the audience?
- Which type of images makes more sense?
- Is the text used more readable or is it too difficult to understand?

The answer to these questions may provide remarkable help to advertisement designers/agencies to attract customers. If advertisements are designed with all these inputs the readability and reach will increase thereby enhancing the product's popularity and sale in the market. See an example of an e-commerce clothing application, consider the two images shown in Figure 1(a) and Figure 1(b) which are used by a brand to promote their product. Figure 1(a) has overmuch text, details, and colors whereas, on the other side, Figure 1(b) has a clear image, less text with fewer details. It is noticed that article viewers in Figure1(b) are much lesser than article viewers in Figure 1(a). The manual observation/ statistical measures are not appropriate to provide an efficient and precise outcome of advertisement popularity [4], [5]. All these questions are completely unanswered due to a lack of technology or analytics [6]. Therefore, the main problem statement is to fill this gap and assist these users to optimize their content and their content strategies [7], [8].

There is a need for a prediction model using handcrafted and auto-generated deep learning advertisement features [9] to get consumers' perception of the advertisement and media content [10], [11], [12]. These consumer perception insights can help the advertising industry to understand the audience preferences and ultimately the most engaging content [13], [14]. Publishers can generate engagements on-demand and can increase the advertisement readership by getting these valuable insights about the advertisement [15]. Therefore, the objective of this work is to prove that the usage of the correct image, correct keywords, and correct Meta features associated with the content can have a great impact on users' engagements- Views, shares, the popularity of any article/blog/advertisement [16]. This technique can avoid the usage of user click fraud in online advertising [17], [18].

This work is an initial impetus to design an analytical model

Anuja Arora working at the Department of Computer Science and Engineering and Information Technology, Jaypee Institute of Information Technology, Noida, India 201310 (e-mail: anuja.arora@jiit.ac.in).

Vikas Hassija is with the School of Computing, National University of Singapore, Sangapore (e-mail: v.hasija@nus.edu.sg).

Shivam bansal is with the Department of Computer Science, National University of Singapore (e-mail: shivam.bansal@u.nus.edu).

Siddharth Yadav is student of the Department of Data Science, North Cap University, India (e-mail: siddharth16csu361@ncuindia.edu).

Vinay Chamola is with the Department of Electrical and Electronics Engineering, BITS-Pilani, Pilani Campus, India 333031 (e-mail: vinay.chamola@pilani.bits-pilani.ac.in).

Amir Hussain is with the School of Computing, Edinburgh Napier University, Scotland, UK (e-mail: A.Hussain@napier.ac.uk).

Digital Object Identifier: XXXXXXXXXXXX

Figure 1: (a) Bright Colour, Enough text, and details (Article Views: 14,0000)(b) Clear Image, less text, and fewer details (Article Views: 90,100)[Data source: Mashable Inc, collected September 2018]

and the objective of the model is to forecast the popularity of online news articles and provide relevant actionable insights to the users. The publishers can rework their strategies to use the right content or images, and post at specific times based on the insights. Ultimately, this can support boosting the popularity of their articles and selling advertisements. With the use of the proposed online advertisement popularity model which is dependent on preemptive analytical insights [19], publishers can garner more traffic on their websites, expect more numbers of clicks on advertisements, more numbers of sales through the platform/website, enhance their existing customer base, add customers with different demographics into their customer base, identify window shoppers and returning visitors or customers, outperform the competitors, improve overall website Alexa rankings, and improve overall Search Engine Optimization rankings.

It simply gives the power to design articles based on real-time feedback and analysis [20]. This proves to be essential in today's world since the number of articles/advertisements developed and published in a day are enormous. Therefore, it is desirable and requires an approach to make an article/advertisement/blog on top.

The organization of the manuscript is detailed herewith. Section II contains the referred research and supporting evidence in terms of studied tools followed by shortcomings found in the existing models. Section III covers the aspects of dataset collection and its exploratory data analysis for understanding. Feature enrichment is detailed in section IV and section V which provide detailed information about techniques used for text feature and image features extraction respectively. The next section (section VI) gives information about all the baseline models and an ensemble model. Section VII discusses experimental settings and the corresponding outcomes. Some enriched features based on experimental findings are discussed in this section. Section VIII finally presents the concluding remarks with some insights into the future scope.

II. Related Research and Supporting Evidence

The US-based organization CoSchedule uses data analytics to score the headlines of the articles [21]. They use several features such as Word Balance, Content Sentiment, Word Meta Features, etc. to evaluate the content of the blog/article headlines. Furthermore, the CoSchedule content analyzer performs modeling and generates a score for the content, headlines, and titles based on the features. This is helpful for the editors to understand the potential outcomes. The sample outcome of the word Balance feature is shown in Figure 2 for a clear understanding.

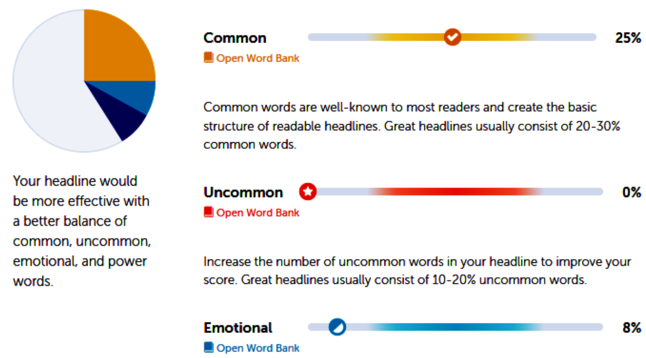

Figure 2: A screenshot of CoSchedule Content Analyzer

One more similar kind of tool is Sharethrough which is a Stanford research project [22]. It has a tool to analyze and provide a quality score (See Figure. 3). The Quality Score is computed with the help of a multivariate linguistic algorithm. This algorithm is built on the concept of behavior Model theory, Sharethrough neuroscience, and advertising research. The multivariate linguistic algorithm considers more than 300 unique variables comprised of EEG data and Natural Language Processing content. These unique variables enable users' native ads to capture attention, expand engagement, and manifest a stronger impression.

Figure 3: A screenshot of Sharethrough Content Analyzer

One of the key shortcomings of these tools is that they do not use Image Data to estimate the expected number of shares on the articles. Additionally, these tools only provide insights about what can be improved, but they do not talk about what are the expected numbers of shares the articles can get. A conventional approach is to consider rating, feedback, sentiments, and demographic information to deal with product prediction problems but all these require some historical data and that is the prime reason for no prediction for a cold start item [23], [24].

Online news popularity prediction and evaluation are studied extensively by researchers. Online news popularity prediction nowadays has become the prime domain to endure consumers and provide inflated targets in form of likes, shares, and leads [25]. Generally, to work in this direction, researchers have used the online news popularity dataset from the UCI repository [26]. In 2015, Ren & Quan worked on the prediction and evaluation of online news popularity [27] and used the same UCI repository dataset. In their work, 20 baseline and text features are used and machine learning algorithms are applied to rank these features. In machine learning algorithm 10 different levels of SVM regression and Random Forest algorithms are used to generate 5-fold cross-validation results. Accuracy varies from 55% - 69% for variant of SVM. Similar research work direction is practised by Namous et. al. in 2018. They also have applied SVM, Random Forest, Neural network [28], Naive Bayes to explore the possibility of accuracy enhancement but achieved accuracy is 65% with optimal parameters [29]. Even, research work specifically for feature enrichment using Boltzmann machine [30], Resnet [31], etc. are performed by researchers. For other application prediction, variants of random forest algorithms are also introduced in recent research work where instead of random subset selection, researchers have proposed a multi-view rand based random forest algorithm [32].

In general, Researchers worked in this same direction and applied numerous machine learning algorithms. Fernandes et. al. [33] is also work in Online News Popularity prediction and validated using ensemble algorithm - Random Forests, AdaBoost; weak learners algorithm- Support Vector Machine, K-Nearest Neighbor, and Naïve Bayes. Performance is evaluated using an overall area under the Receiver Operating Characteristic (ROC) curve was 73% and acceptable discrimination and accuracy was around 67% [33].

Bhargava et. al. used KNN, Random, Naive Bayes, and Neural to predict news popularity. The result analysis is done using correlation analysis, particle swarm optimization, and principal component analysis. Result analysis is performed in a different way than previously done work. The result shows that SVM and naive outperform with correlation algorithm and KNN, on the other end neural outperforms PSO [34] [35].

One recently published research paper measured the amount of attention to news articles. The author primarily focused on dimensionality reduction and how irrelevant attributes deteriorate the learning algorithm performance while predicting. A new feature selection strategy is proposed in this research paper which enriches features/ attributes using Cellucci-Mutual Information-Based feature selection technique [36]. The author claimed that the approach can extract the most important features and can accurately predict news popularity [36]. The existing researcher worked on fusing and mining opinions from reviews. The prime point of discussion is the correlation/degree of relevance between reviews and their associated news article in published research work of the last few years [37]. Feature-based customers' effective response has been extracted by wang et. al., their work uses seven affective attributes/ features, and heuristic deep learning is modeled to get the relationship between review texts and customers' response [1]. Text features are extracted to compare unrelated reviews and articles. Article reputation is measured based on a fused principal opinion set i.e. opinions of similar and same attitudes and preferences [38]. Some researchers focused their work on images and video-based news popularity [39] as well. Various techniques are used to predict popularity by extracting the image and video features such as Fourier transform [40], image visual understanding-based features such as contrast [41], surveying [41], and Topic and sentiments for image Ad [41], Competitive correlation Keywords used in advertisements [42]. Patents have also been published to predict popularity using Images. In this patent global feature, regional features, and image segmentation is used and hidden unit response is extracted using a restricted Boltzmann machine. Finally, search results are ranked based on these features based on image quality [43]. In 2017, ACM Multimedia introduced a challenge to predict popularity of photos on Social media Prediction Task-1 (SMP-T1). Several teams participated in the challenge and out of all One research work presented in this challenge was by Wang et. al. [44] on collected Flicker Dataset. In this research work, gradient boosting regression tree (GBRT) used ridge regression for prediction score generation by modeling the linear regression [28] [45] in between features. Overall 10 features- user id (distinct key), post date, number of comments, hashtags, length of title, length of description, tag count, average view, group count, and average member count are used to predict population using Ridge Regression method [44]. Performance is measured using Mean Absolute Error (MAE) and Mean Squared Error (MSE) which was 1.1059 and 2.1767 respectively. Liao et. al. [46] research work used recurrent neural network (RNN) and Convolution neural network (CNN) for the deep fusion of temporal process and content features model (DFTC). The dataset which is used to validate news popularity is from WeChat. The temporal process is modeled using RNN to capture long-term growth trend of popularity and for short-term features attention, CNN is used. For content features Hierarchical Attention Network [47] and embedding techniques are used for capturing text and Meta features. Competitive keywords based on online advertising is also been validated to improve search engine advertising performance [42].

Recent studies in this direction use text, image, and video characteristics. Gkikas et. al. used text characteristics - readability indices, text length, and hashtags to get to know the relationships of post content with consumer engagement and brand awareness [48]. A comparison of visual and linguistic metadiscourse is performed to get an insight into enticing more customers into buying products [49]. The outcome of this study is more toward theoretical justification with manual parameter extraction and manual inspection. To prove the importance and efficiency of automatic feature generation for advertising, an automatic code campaign for political advertising videos is proposed which shows a comparison of machine learning code and human code for various audio and image features [50]. Experiments present improvement in the efficiency and scope of campaign advertisement research with the help of machine learning. Still, an empirical and prototype computational model for advertisement popularity prediction based on these machine learning features is missing in the published literature that will undoubtedly be a great support to the advertising industry. The computational models that can provide text, image, and video features based on advertisement popularity prediction can provide generic and genuine support to advertising agencies of different domains such as education, healthcare [51], supply chain management [9], travel tourism [52], etc. For any domain, enticing and engaging the customer with products is done using published media content (text, image, or video) and prediction of popularity by finding an association of content with consumer engagement is the viable and only solution for advertising of any domain.

In the context of the dataset, we have taken the Mashable inc dataset of UCI repository but we have not taken that as it is we have extended both text feature dataset [53] and image feature dataset [54]. Various language techniques and scores are used to upgrade the text feature dataset of all the articles' content [55]. Similarly, ResNet [56] is used to fetch Image features that are high in amount so PCA [57] is used for dimensionality reduction of ResNet provided approximately 10,000 features.

Generally, the foregoing research studies focus on regression models and machine learning models usage in various domains such as healthcare, privacy and safety [58], e-commerce [59], medicines to detect, identify, and predict influencing factors, adverse effects, adverse factors, patterns, etc [60]. Iglesias et. al. used evolving fuzzy systems to classify real-time news articles into various topics [28]. In this work, news articles are categorized into various topic areas and this classification has been done utilizing the text content of the article. Work is performed on continuous, time-varying news article data as well to handle data streams, a web news mining framework is proposed by making use of evolving Type-2 classifier [37].

These found research gaps in studied literature framed the direction of research work presented in this manuscript. The existing work uses textual features or image features individually which makes the solution approach inconsistent. Henceforth, there is a need for a robust solution approach for variety of features driven news popularity prediction. The proposed solution approach is drafted in a hybrid manner i.e. added combination of baseline features (social context features), textual features using natural language processing, and Image features using an image-based deep learning model. Even though a similar sort of work has been done separately on other application areas but just a few features have been used for a particular content type (text, image). As per our knowledge, readability, a textual feature have not been used for popularity prediction of an article, and extensive image features extracted using Resnet deep learning architecture are also not been used for online news popularity. Another novelty of work is learning the importance of feature or assignment of weight to features based on mapping of feature and news popularity index. All three categories of features (Baseline, Textual, and Image) are highlighted based on their importance. At last, Several baselines and ensemble learners in combination with features categories are used for predicting news popularity using identified features and corresponding feature weight.

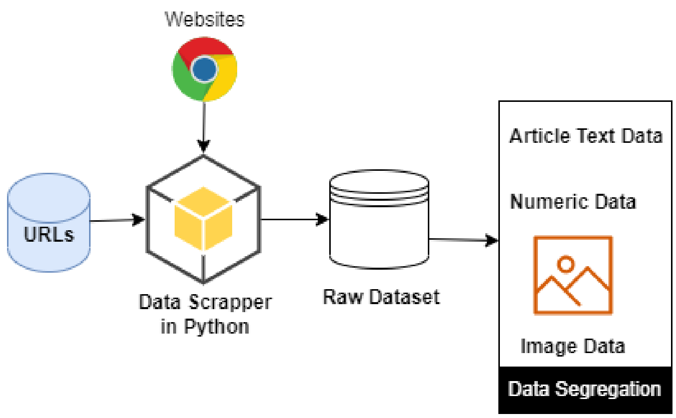

Figure 4: Data Collection Process

III. Data Collection and its descriptive analysis

After several research efforts, the Mashable inc website is used to extract the details of articles. 40,000 articles details along with their final number of shares have been published in total from 2013 to 2015. The final number of shares acts as the target variable, and the use of past data makes more sense as essentially all the articles are now in saturation and their shares will not change. So, as shown in Figure 4, Title, Shares, Main article image, Article text, and Article tags are present on Mashable. inc for all the articles. The data set is a public dataset and it is available in the UCI Machine learning repository [26]. We have used only two fields from the complete dataset-Articles, i.e., URL, and Number of shares. Rest all the features are generated by scraping of data. Three kinds of data have been scrapped from the articles' URL pages. The extracted data categories are Baseline data (Views, shares, number of href, etc), text data, and image data. The data collection process is shown in Figure 4.

A. Baseline Data

The baseline dataset includes the features and information that was published in the public dataset. These details include article URL, article published Date, number of reference articles, images count, videos count, and shares count. The shares count is considered as the target variable. This baseline data has been collected for all 40,000 published articles by Mashable Inc.

B. Text Data

Further, to validate news popularity based on its content and associated image, we decided to enrich the data set by adding some key features. The first form of feature addition includes the text data information that has been crawled from every article using the article URL. The web crawling steps have been performed using python scripts. The script visits every article page and retrieves the details from the article and saves it. The added text key features are article title, article text (content), author name, and article tags. These features are added alongside the baseline dataset.

C. Image Data Extraction

The next set of features that have been added to the data set is the Image used in the article. The images of all the articles have been crawled using the Image URL obtained from the previous step. So, the baseline dataset has been enriched with new features that include the text data content and the images used in the article.

D. Data Preprocessing

Data preprocessing is required for both baseline and text data. The text data consists of raw content such as URLs, Stop words, Expressions, etc. Henceforth, data cleaning is mandatory and the processes that are applied to clean data are

- **Escaping HTML characters:** Removal of HTML tags those are embedded in the original data such as < > &
- **Apostrophe Lookup:** Any word sense disambiguation in text needs to be avoided, In our work, all the apostrophes are converted into standard lexicons using lookup tables.
- **Removal of Stop-words:** In-text cleaning, we removed the commonly occurring words (stop-words)
- **Removal of Punctuation:** Symbols and special characters were removed
- **Removal of URLs in the text:** URLs and hyperlinks in text data were removed.

Further, preprocessing is performed on baseline data. The following steps are performed on all the baseline features present in the dataset:

- **Min-Max Normalization:** All the independent Variables were normalized using the equation

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

- **Logarithmic Transformation of Target Variable:** We converted the target variable to logarithmic to make the target variable follow the normal distribution.

E. Exploratory Data Analysis

The exploratory analysis helps us to learn more about the target variables (number of shares), and some of the key features such as - Images count, Videos count, and the number of reference articles [61]. The exploratory data analysis is performed on the dataset to understand the insights of features such as how the features are spread across every article? In the collected data shares are the dependent target variable and all other features- Number of Images, Number of Videos, Number of Hrefs, Top Tags Used are dependent features. The distribution of Dependent target variable - shares of articles is shown in Figure 5.

Other baseline independent variables- article image distribution, article video distribution, and article Href distribution are shown in figure 6, 7,and 8 respectively.

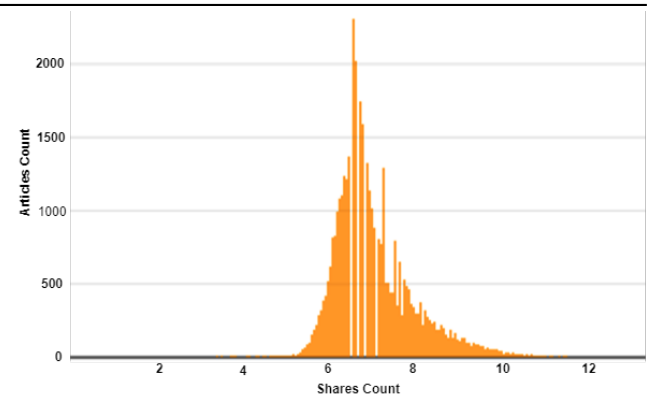

Figure 5: Distribution- shares of articles, X Axis: Shares Count, Y Axis: Articles Count

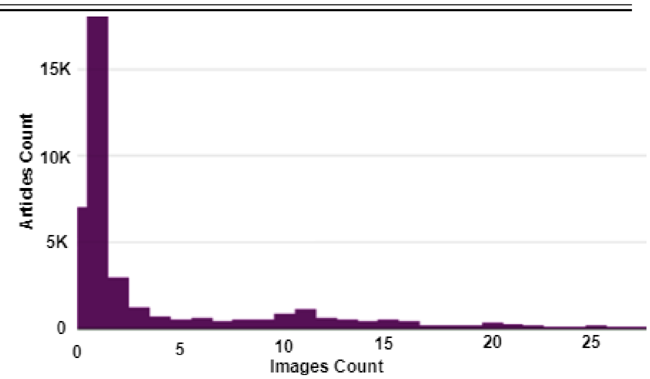

Figure 6: Distribution- Images in articles

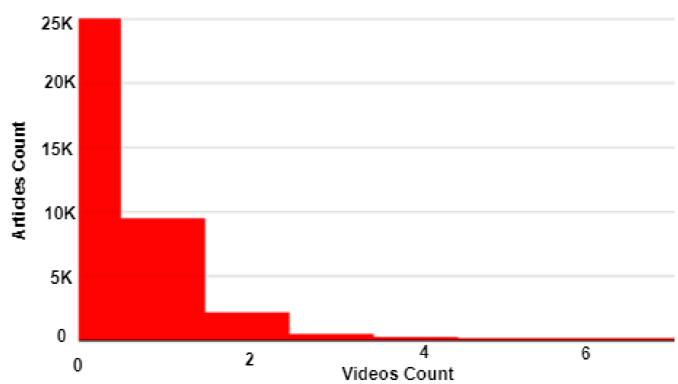

Figure 7: Distribution - videos used in articles

IV. Text Feature Engineering

Five kinds of features are obtained from the text data by applying the corresponding techniques. The details of text feature engineering and the workflow are shown in Figure 9. The five incurred features are- Readability [62], Meta Text Feature, Sentiment Features, Dictionary features, and TF-IDF Features [60]. The process of obtaining these features is detailed in further subsections.

A. Readability Features

Readability features are used to evaluate the readability of text. Readability is defined based on the count of syllables,

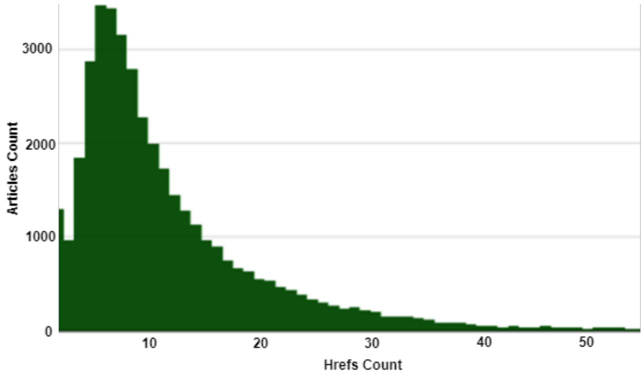

Figure 8: Distribution - Href used in articles

words, and sentences exist in an article. Readability tests considered as a viable and capable alternative to conduct an actual statistical survey of readers of the subject content. A useful readability test protocol will give a rough indication to measure readability of a particular work. For a high volume of content the accuracy of readability increases. Based on the heuristic, an article which is having high readability will get a higher number of views and shares. The features which have been generated to measure the readability of articles are Flesh Reading ease, Flesch- Kincaid grade level, SMOG Index, Coleman-Liau Index, and Lix.

1) The Flesch Reading Ease formula [63]: This formula is used to evaluate the readability ease of English language written content. The output, Readability Ease (RE) values lie in the range of 0-100. A higher score indicates the article is easy to read and the average output of RE is between 6-70. The mathematical equation used to compute RE is depicted in equation (2)

$$\mathrm{RE} = 206.835 - (1.015 * \mathrm{ASL}) - (84.6 * \mathrm{ASW}) \tag{2}$$

Where, RE = Readability Ease, ASL = Average Sentence Length, and ASW = Average number of syllables per word.

2) The Flesch-Kincaid Grade Level [64] : This measure is used to assign a score to the text based on its difficulty level. The level indicates that the average student in a spcific grade level can read the text. The higher grade level indicates difficult to read text. The mathematical equation is used to compute the Flesch-Kincaid Grade Level (F-KGL) is in equation (3)

$$\mathrm{F-KGL} = (0.39 * \mathrm{ASL}) + (11.8 * \mathrm{ASW}) - 15.59 \tag{3}$$

3) The SMOG Index [65]: This test is used to measure the readability based on average persons can read the text. A higher value means the text is difficult to read. SMOG Index is computed by counting the number of sentences as well as the count of polysyllabic words in these sentences. SMOG Index is calculated with equation (4).

$$\text{SMOG Index} = 1.0430 * \sqrt{\text{number of polysyllables} * \frac{30}{\text{number of }_{\text{sentences}}}} + 3.1291 \tag{4}$$

4) The Coleman-Liau Index [66]: This test relies on characters contrary to syllables per word and sentence length. This measure used to compute Coleman –Liau Index (CLI) score is written in equation (5).

$$\mathrm{CLI} = 0.0588\ \mathrm{L} - 0.296\ \mathrm{S} - 15.8 \tag{5}$$

Where L is the average number of letters per hundred words and S is the average sentence per hundred words. red

5) LIX Score [67]: This test is used to calculate readability measures based on the difficulty of reading a foreign text. The LIX score is computed using equation (6)

$$\mathrm{LIX} = \mathrm{A/B} + (\mathrm{C} \times 100)/\mathrm{A} \tag{6}$$

Where, A is the number of words, B is the number of periods, and C is the number of long words

B. Text Meta Features

Numbers of extra text-based features are also generated. These features have already been used in studies by researchers for improving text classification models performance. The text meta-features computed in this research work are:

- Word Count of the document;
- Character Count of the words in a document;
- Average Word Density of words in the documents;
- Unique Clean Words.

C. Sentiment Features

The quantification of sentiments associated with the article and its title are used as two features. To measure this TextBlob package in Python has been used. TextBlob is a rule-based sentiment analyzer and supports in measuring polarity and subjectivity of article. A research article published by Chaturvedi et. al. discusses hand-crafted features and atomized models for subjectivity and polarity detection [68]. Polarity is only used as subjectivity quantification is not required in case of news popularity. Polarity is used to quantify emotions expressed in a sentence. The polarity is the sentiment ranging from -1 to +1. The negative value of polarity interprets a negative emotion and positive value interprets a positive emotion and 0 values means neutral emotion. It is noticed that articles that share positive content are likely to garner the number of shares.

D. Dictionary Based Features

To make a post popular emotional and powerful words are also crucial and significant. These emotional and powerful words that can lead to more shares are also explored by the coschedule.com team. Therefore, Dictionary-based features are added as additional features. These dictionary-based features are added to measure the impact of power words and emotional words in the article data. The Power and Emotional word dictionary are obtained from coshcedule.com.

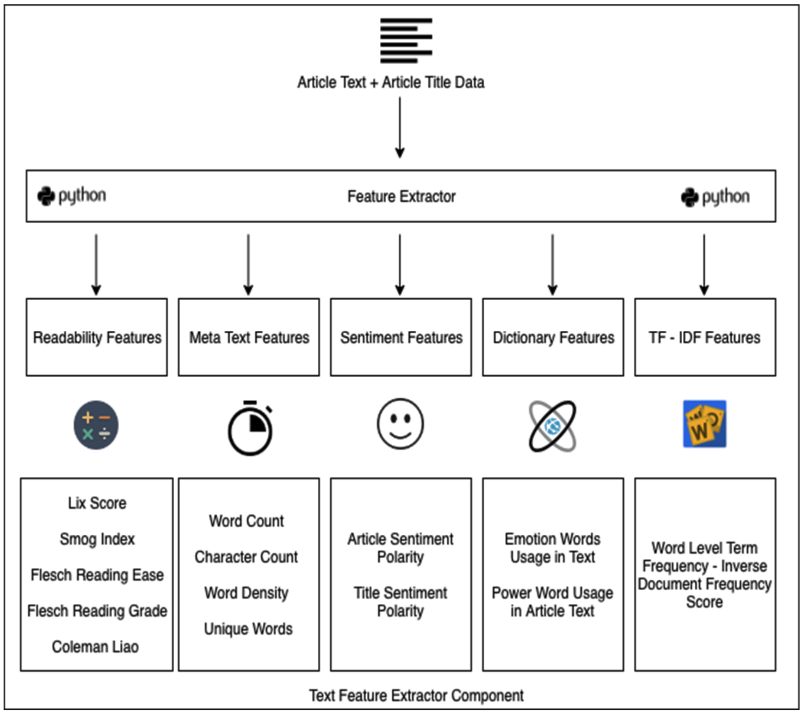

Figure 9: Workflow of Text Feature Extraction

V. Image Feature Engineering

Several State-of-art CNN architectures- AlexNet, LeNet, VGG, LeNet are used to provide the solution to the most groundbreaking computer vision and deep learning area. All these architectures are going deeper and deeper using the number of layers. AlexNet has 5 convolution layer, VGG has 16 and 19 convolution layer, GoogleNet has 22 layers. Deep Neural systems do not train systems effectively just by increasing stacking layers. This adds a vanishing gradient problem so ResNet is introduced. Deep Residual Network is used to train thousands of layers to achieve compelling performance. To generate image features ResNets uses skip connection and jump over a single layer. The concept behind the skip layer is basically to avoid the problem of vanishing gradients by reusing activations from a previous layer until the layer next to the current one has learned its weights. In the ResNet initial phase, the neural network collapses into fewer layers to make it easy to learn and layers expand step-by-step to learn more about the feature space. When layers are expanded it contains many features and learns faster. In this work to learn image features of advertisement, a pre-trained ResNets model is used to generate the low-dimensional features from the dataset. Resnet model works on the concept of skip-connections in which the activations and the inputs are added together and propagated forward. This helps to avoid the vanishing gradient descent problem. Therefore, a pre-trained ResNet model is used and deactivated the last layer to obtain the final output. Figure 10 showcases the Overall Image feature generation process. The pre-trained ResNet50 model is used which generates low-level image features. It generates 10,035 dimensions of low-level features for our advertisement image dataset. Further, Principle Component Analysis (PCA) is applied to reduce dimensions [69]. PCA is a dimensionality reduction technique that provides a lower dimensionality approximation of the original data while preserving as much variability as possible. It uses an orthogonal transformation to transform a set of correlated features into a set of linearly uncorrelated variables, those are known as principal components. The largest possible variance exist in first principal component and each succeeding component in turn has the highest possible variance with the constraint that it is orthogonal to the preceding components. usually after data normalization, PCA is performed by eigenvalue decomposition or singular value decomposition of a data matrix. Therefore, It is used to reduce the dimensions of image features from 10,035 into 100 principal components [68].

VI. Baseline Models

Baseline features, Text-based features, and image-based features have been extracted using various techniques as discussed in section IV and section V. Furthermore, to predict shares count four baseline regression models have been used. The models which are used to predict share counts are- Linear Regression, K-Nearest Neighbour Regression, Random Forest, and Extreme Gradient Boosting model. A combination of the above given three categories of features is used to predict shares using these four baseline models. Hence, the above 4 baseline models are trained on the three different dataset compilations: Baseline data, Baseline data with text features, and Baseline Data with text features and image features. The complete model architecture is shown in Figure 11 which shows that all the baseline models have been applied initially on Baseline features. Then, all four models are applied to the combination of Baseline and text features and finally, Baseline, Text, and Image features. A Total of 12 models, those are a combination of features set and baseline models are used to train the model and test the outcome. These models are tested and validated on different dataset/ feature sets to compare and judge how the accuracy is improved on adding the new features to the baseline data. Twelve models details are enlisted in Table I. The baseline models which are used in order to build share prediction systems are discussed in subsections.

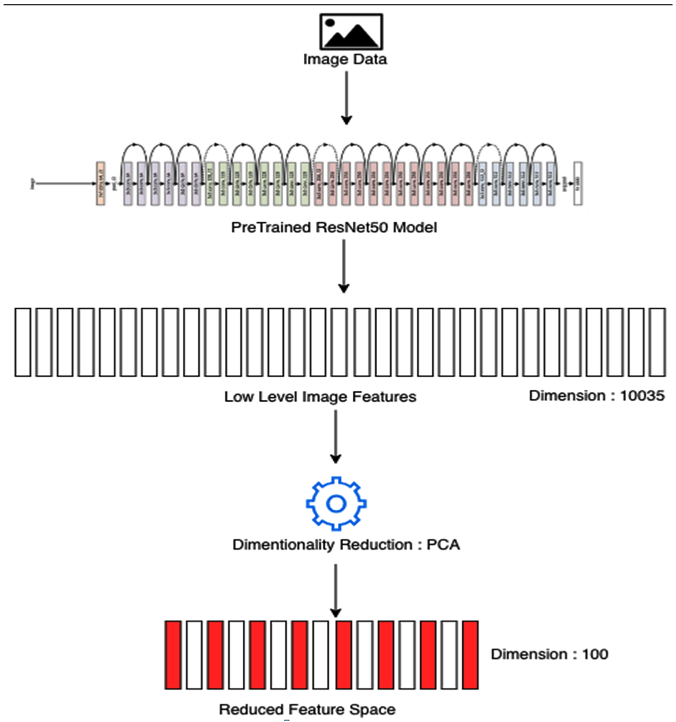

Figure 10: Image Feature Engineering Process

A. Linear Regression Model

Linear regression is used to predict the value of an outcome variable Y based on one or more input predictor variables X. The aim is to establish a linear relationship between the predictor variable(s) and the response variable. The Linear regression model is chosen so that the model can predict the linearity of the factor variables and the predictor variables. “lm" function of R language is used to train the regression model.

B. K-Nearest Neighbor Regression

K-nearest neighbor is used to predict the value of outcome variable Y based on the k-nearest neighbor Y. The closeness is based on the Euclidean distance of other variables between them. The value of Y will be the average of the Y values from its k-nearest neighbors. The reason we chose the KNN model is that it gives a decent level of prediction and is easy to interpret the output. Even, the time required for the model is less. To Train the KNN model, the “caret" package is used with 5-fold cross-validation. Further, the K value is tuned from 5 to 43 with all the odd numbers, and finally K= 35 for the best result.

Table I: Model Description.

	Model Type	**Feature Set**
Model 1	Linear Regression	Baseline Data
Model 2	KNN	Baseline Data
Model 3	Random Forest	Baseline Data
Model 4	XGBoost	Baseline Data
Model 5	Linear Regression	Baseline Data and Text Features
Model 6	KNN	Baseline Data and Text Features
Model 7	Random Forest	Baseline Data and Text Features
Model 8	XGBoost	Baseline Data and Text Features
Model 9	Linear Regression	Baseline Data along with Text and Image Features
Model 10	KNN	Baseline Data along with Text and Image Features
Model 11	Random Forest	Baseline Data along with Text and Image Features
Model 12	XGBoost	Baseline Data along with Text and Image Features

C. Random Forest-based Regression

Random forest is the assembly outcome of multiple decision trees by merging these decision trees to get a more accurate and stable prediction. Undoubtedly, It has better predictive accuracy as compared to a single decision tree and works well with default parameters settings. The Random Forest is applied to the baseline data which essentially is a regression problem. Random Forest model is created by tuning ‘mtry’ and n tree values using ‘caret package’ which essentially specifies the level of trees that can be used for splitting and the number of iterations to run with. Even, 5 fold cross-validation to find the best mtry. The final value of mtry $= 2$.

D. Extreme Gradient Boosting: XGBoost

XGBoost is an execution of the famous gradient boosting algorithm. XGBoost contains a series of models where each new model is built to predict the previous model’s error. To make a prediction using xgboost, we will need to add up the prediction of all the models. XGBoost is known for its fastness speed and accurate predictive power. We trained the XGBoost model by tuning all the tuneable parameters *max_depth*, *min_child_weight*, *gamma*, *subsample*, *colsample_bytree*, *nrounds*, *etc* in “caret". Even, 5-fold cross-validation is applied to find the best parameter sets.

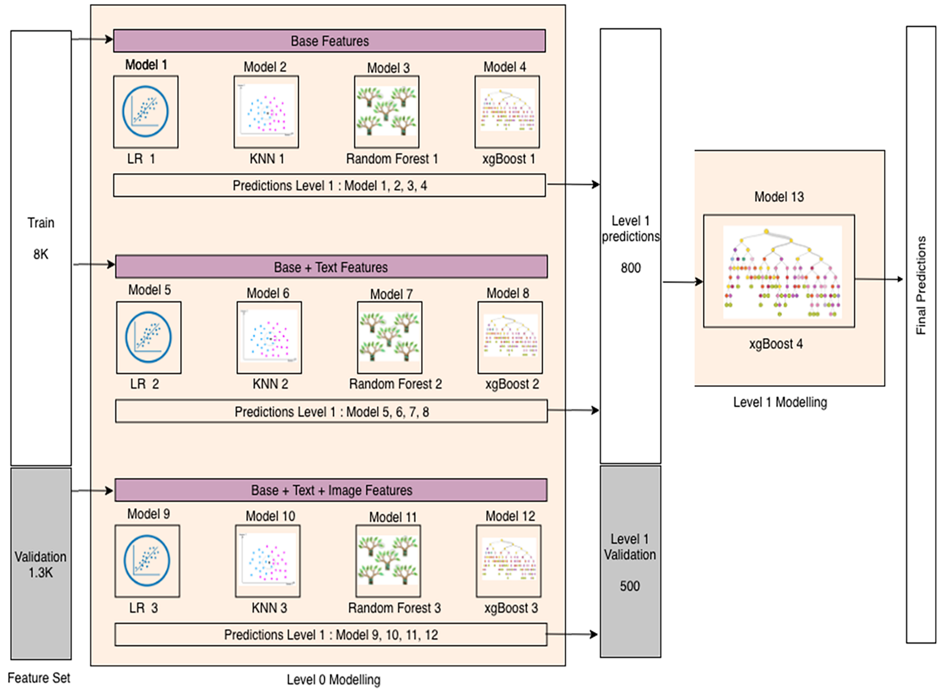

Figure 11: Model Architecture Setup

E. Ensemble Modelling: Stacking Strategy

An ensemble model is an assembly or composition of various machine learning techniques to reduce generalization errors in machine learning tasks. These methods use multiple learning models to achieve better predictive performance in contrast to any of the constituent learning models alone. The stacking technique is employed as a method of ensemble modeling. The fundamental concept behind this is to use a pool of weak learners (base predictors) and then use ensemble predictors to combine the base predictors' outcomes. The base predictors in our Stacking model are the prediction values of the 12 baseline models trained on the training sets. The Shares predictions of baseline Level1 models are now featured for Level 2. In level 2 XGBoost is used as a model. XGBoost as a Meta learner is significantly more accurate compared to many other base learner outcomes. Additionally, the model can predict effects on the test set for the final evaluation metric.

VII. Experimental Outcome: Data Insights, Analysis, and Features Importance

Several insights and outcomes are displayed in this section. These insights are required to understand the complete concept behind online popularity and its associated baseline, text, and image features. Detailed and clear results are depicted in the further divided five subsections.

A. Exploratory Data Analysis (EDA): Text Features

Exploratory data analysis is shown through the mapping of text features with shares in the article. The size of the bubble shows the shares count concerning readability features. Figure 12 shows the article shares distribution according to computed the SMOG Index and the Coleman Liau Index. Figure 13 is about the number of shares distribution based on the article title and article text polarity. The third shares distribution shown in Figure 14 constitutes shares distribution based on Emotional and Power words used in the articles.

Articles Shares corresponding to varying categories of features (as shown in figures) shows the equal distribution of features of considered articles. This exploratory data analysis is needed to validate that the generated features set are not suffered by class imbalance problem which is clearly visible through these distribution graphs.

B. Text Features Correlations

Several textual features are extracted through diverse natural language processing techniques (discussed in detail in Section IV). To explore correlation, a Pearson correlation mapping of every feature with all other remaining features is presented with the help of heat map in Figure 15. In the heat map,
- Positive values represent the positive correlation in two features and are shown by pink colors, such as, the 'word density' feature and the readability feature 'Lix' are having a positive correlation with the value of +0.25.
- Negative values represent the negative correlation in two features and are shown by blue colors, such as, 'word density' feature and readability feature 'Flesch reading ease' is having negative correlation with the value of -0.75.
- Neutral i.e. zero represent the independent features i.e. two feature is neither positively nor negatively correlated which is shown by grey color.

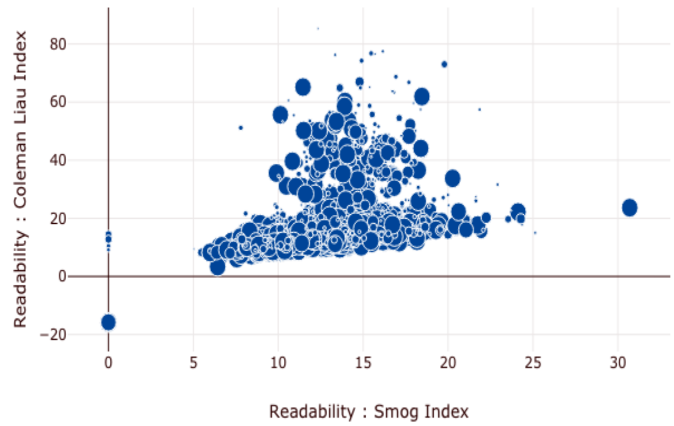

Figure 12: Articles Shares according to readability features

Figure 13: EDA:Articles Shares Corresponding to Sentiment Features

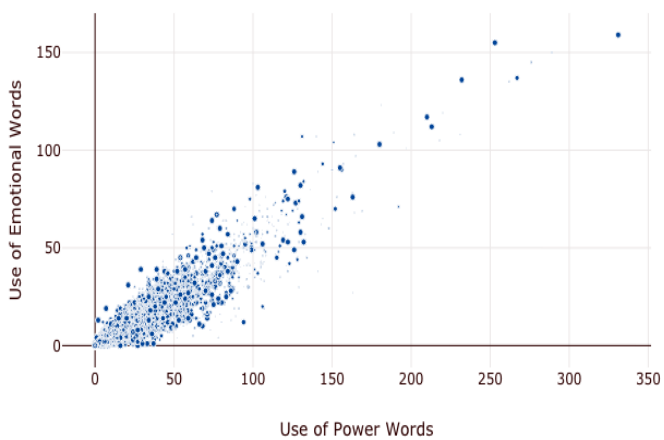

Figure 14: EDA:Articles Shares Corresponding to Dictionary Features

Figure 15 is self-explanatory heat map and clearly depicts the positive and negative relationship in text features.

C. Dataset Split and Evaluation Metric

The data set splits into the Train, Validation, and Test set for training and evaluating the prediction accuracy on

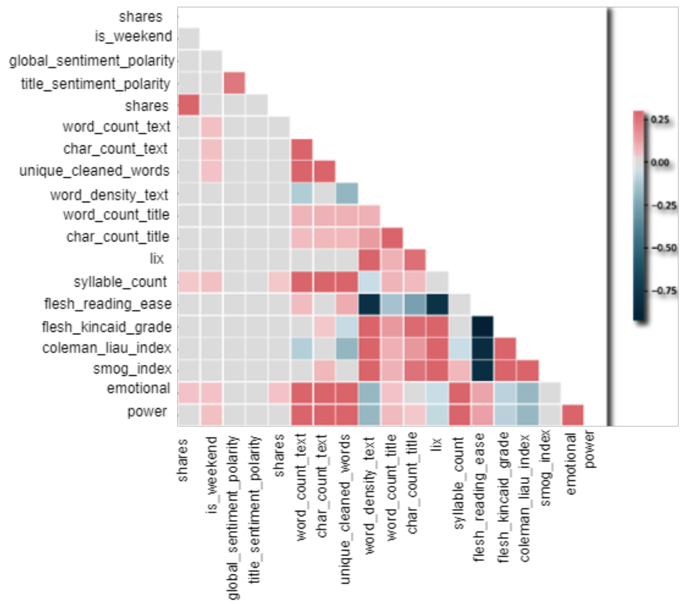

Figure 15: Correlation in Text Features

varying features – Baseline features, Text features, and Image features. The sample size is taken using the Image dataset as a reference and then split into two for train and validation. The validation set is again split into validation and test set and which will be used for accurate prediction. The same applies to the other data set. The training data contains 8k articles for baseline and text data and 5k for image data. Validation and Test set are the same for all, respectively 800 and 500 articles. Root Mean Squared Logarithmic Error (RMSLE) is used for performance evaluation and validation of results. RMSLE measure is used to find out the error but calculate the difference between the values predicted by the machine learning model and the actual values. In this research work, the outcome of the model is in the form of the Number of shares i.e. how many numbers of shares will be achieved by a specific article. The outcome can't be matched with exact/ precise numbers. Hence, an evaluation measure is required which will sanction the difference in shares for a threshold and RMSLE is the best measure in that sense. RMSLE is generally the best-considered performance measure when the system doesn't want to penalize huge differences in the predicted and the actual values i.e in cases where both predicted and true values are huge numbers. Hence, RMSLE is computed using equation (7)

$$RMSLE = \sqrt{\frac{1}{N}\sum_{i=0}^{N}\left(\log\left(y_i\right) - \log\left(y_i\right)\right)^2} \tag{7}$$

D. Feature Engineering Outcome

Importance of Features as a resultant of XGBoost model is presented in Figure 16, Figure 17, and Figure 18 for all three categories- Baseline, Text, and Image features respectively. Figure 16 shows the resultant importance of baseline features where the number of images, Category of the article (World

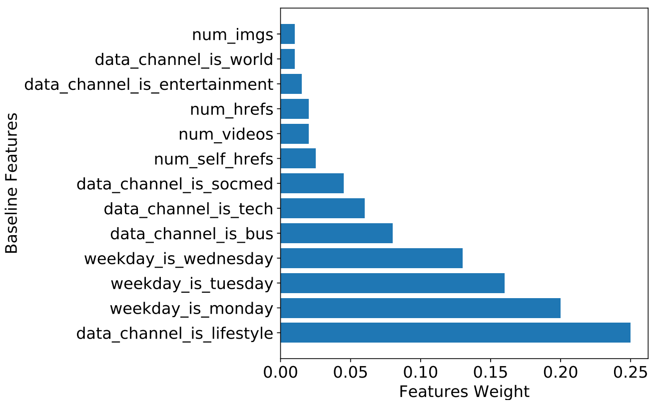

Figure 16: XGBoost Based Category wise Features Importance Distribution of Baseline Feature Importance

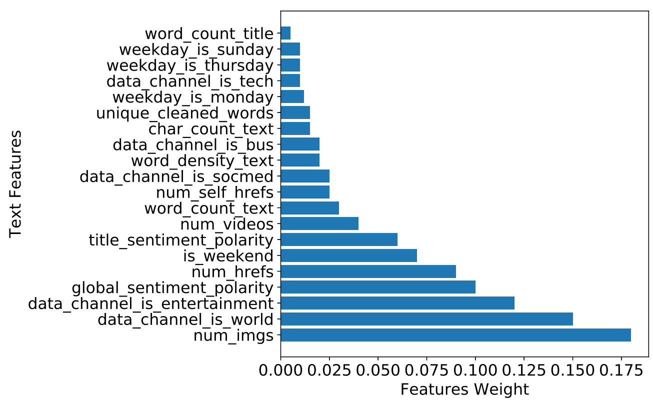

Figure 17: XGBoost Based Category wise Features Importance Distribution of Text Feature Importance

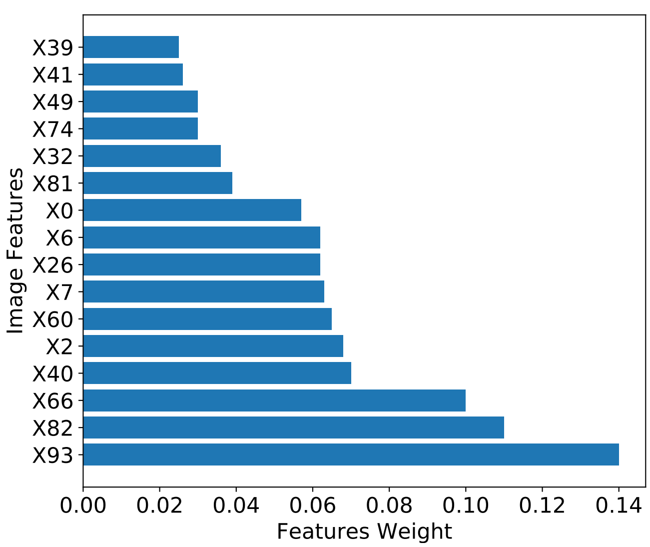

Figure 18: XGBoost Based Category wise Features Importance Distribution of Image Feature Importance

and Entertainment), number of reference articles, the number of self-reference article links, and the number of videos are found to be having higher importance value as compared to other features. Figure 17 shows the combination of baseline and text features importance ranking order after applying XGBoost. Out of all text features Global Sentiment polarity, Word Density text, and Word count, and title sentiment along with other important features from the baseline data are all found to be having higher values of importance. Finally, Figure 18 shows the importance of ranking order of all three categories of features. Out of all, Image features are having maximum importance, hence coming in the top list of the feature importance ranking. Few Image features are having extremely higher significance as compared to the other features in the XGBoost model. The ranking order of features is clearly shown in figure 18. The key finding of this feature engineering task is that image features are the most important features [70] i.e., all the top and important features are image features.

E. Features Coefficients and Significance

Table II shows the coefficient/ Estimate of the independent variable in the linear regression equation. The table presents- 1) P-value that lies in 0-1 range, 2) the significance level to determine the statistical significance of each explanatory variable on the explained variable. It is clear that lower the P-value, the higher is the significance of the variable. In table, significance level is showcased by the number of the asterisk symbol ($*$), a variable can have a maximum of 3 asterisks and a minimum of 0. Further, the level of significance is depicted by a full stop (.). It is observable that in the linear model on text features and baseline features, the title sentiment polarity has a strong significance in the model and has $**$ significance.

This implies that the more positive the title of the article, the more shares it will get. Furthermore, the significance of image features for principal components 0, 1, 37, 39, 41, and 93 is high. This indicates that image features play an important role in deciding the number of shares an article can get. Additionally, the features such as the number of hrefs used in the article, the number of images used in the article are crucial and the most significant features to the model as well. Figure 19 and Figure 20 infers that the number of shares increases as more number of images and the reference link is added to the article. Figure 21 and Figure 22 image feature partial dependence plot shows the number of shares increases due to image features showing that the images have a high tendency to increase shares.

Root Mean Squared Logarithmic Error (RMSLE) is used to validate the number of shares predicted by different models and its validity corresponding to actual shares. Table III presents the outcome of the train, validation, and test set for all the twelve models detailed in section 6 and Table I. RMSLE error for all models on training, validation, and test set gives test set error in the range of 0.36 to 0.386 which is itself an effective outcome. The proposed ensemble algorithm is able to achieve 0.321, 0.337, and 0.351 RMSLE errors for training, validation, and test set respectively(Refer Table III). All other models the giving more relative error as compared to the proposed ensemble model for online article popularity prediction.

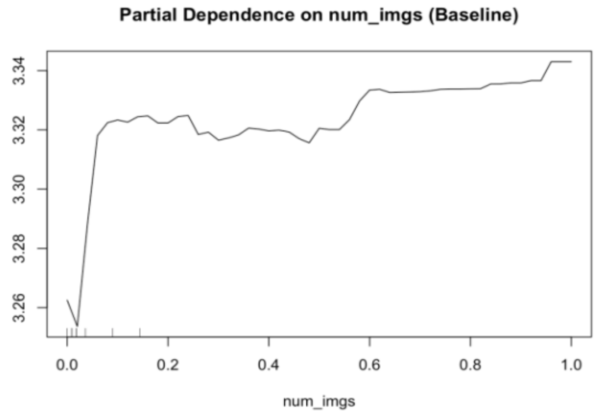

Figure 19: Partial dependence on number of images

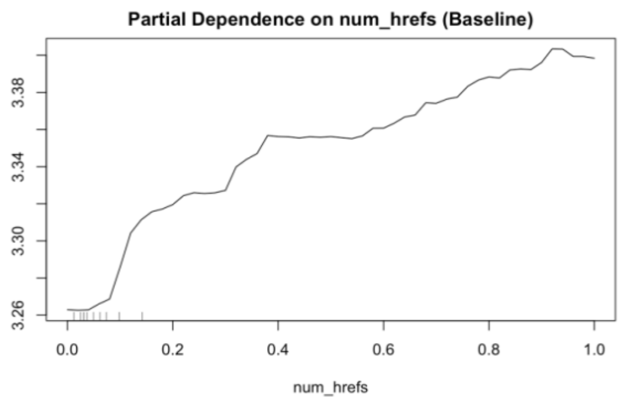

Figure 20: Partial dependence on number of Hrefs

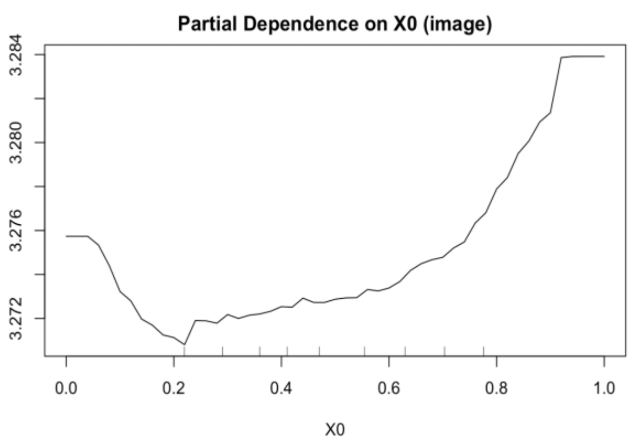

Figure 21: Partial dependence on X0 Image features

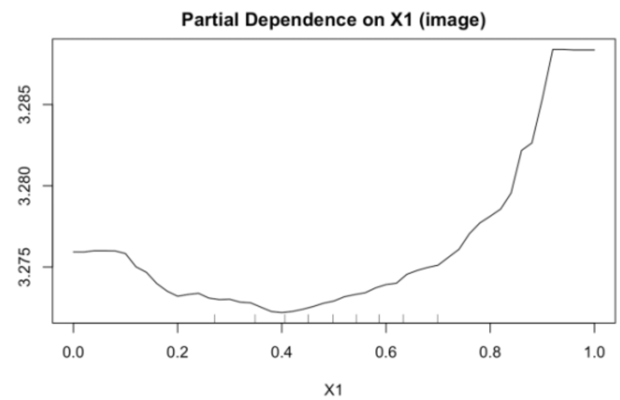

Figure 22: Partial dependence on X1 Image Features

VIII. Experimental Findings

Over recent years, content on social media has greatly facilitated the way marketers, brands, and industries can communicate and convince users. Subsequently, the article context, social context [71], and image context have a great role in influencing customers.

In this research paper, news popularity is predicted based on all these three context information. Baseline features come under the category of social context, text features are under the category of article context, and similar way image context contains image features. Therefore, the enriched feature set in all three directions is the prime cause in finding news popularity efficiently. This section discusses observations and key findings regarding features that are perceived behind the curtain.

A. Impact of Text Features on news popularity prediction

The text features seemed to have higher variable importance as compared to the baseline features. The outcome of the evaluation metric indicates that text features have decreased share prediction and enhanced its performance slightly by 0.01 but the significance of text features is high. Readability features hold 7 ranks out of 15 important features. Specifically, LIX Index, Syllable Count, and Flesch reading ease are is at a higher importance level. The effect of dictionary features is marginal and power words contain higher importance as compared to emotional words within used dictionary features. The effect of Meta features is even lesser than dictionary words. Out of all text features, sentiment features of both title and article are the most significant features. Title Sentiment Polarity is the most influential text feature and has proven its significance and usefulness towards shares prediction.

B. Impact of Image Features on news popularity prediction

It is observable from feature importance distribution that image features are the most significant features and tend to have higher variable importance. Adding them to the baseline and text features has shown vigilant improvement in the news popularity prediction RMSLE evaluation result. The RMSLE value was improved by 0.03 after adding the image features. Intuitively, images have the advantage of delivering the article message effectively. The image feature with a specific theme or color will be very effective since it conveys the mood of the article. This has also been proven by the partial dependence plot of base models. Additionally, the number of images also improves the performance significantly which signifies that adding more images/videos can greatly improve the news popularity. The low-level features such as edge, color, items in an image, etc. all contain their position in the top importance list during news popularity computation. This signifies that even the minor details present in the images can impact the model's performance and news popularity. The overall actionable insights for a publisher is that the article publisher should increase the number of images, use of high-quality images (less blurry, non-textual), number of videos, amount of sentiment in the title, amount of sentiment in content, and more powerful and emotional words. On the other end, the article publisher should decrease the number of longer words in the content, the amount of multi-topic discussion in an article, the number of negative words.

Table II: Feature Coefficient and significant level Table

	Estimate	**Std Error**	**t Value**	**Pr(>|t|)**	**Significance**
Title Sentiment Polarity	0.0943256	0.0306562	3.077	0.0021	**
Char Count Title	-0.0870152	0.0468147	-1.859	0.0631	.
Intercept	3.2957439	0.4665926	7.063	1.84E-12	***
X0	0.0701925	0.025423	2.761	0.00578	**
X1	0.054895	0.0318134	1.726	0.08449	.
X37	0.0642818	0.0388685	1.654	0.09822	.
X39	-0.0956658	0.0442725	-2.161	0.03075	*
X41	-0.0891953	0.0426206	-2.093	0.03642	*
X93	-0.0752779	0.043266	-1.74	0.08194	.
num_href	0.464868	0.0795129	5.846	5.33E-09	***
num_self href	-0.1309966	0.0611654	-2.142	0.03227	*
num_imgs	0.2519118	0.0719822	3.5	0.00047	***

Table III: RMSLE based news popularity prediction result validation

	RMSLE Train	**RMSLE Validation**	**RMSLE Test**
Model 1	0.367	0.381	0.386
Model 2	0.358	0.385	0.382
Model 3	0.353	0.38	0.386
Model 4	0.359	0.371	0.374
Model 5	0.366	0.372	0.376
Model 6	0.363	0.377	0.379
Model 7	0.159	0.373	0.371
Model 8	0.350	0.369	0.366
Model 9	0.362	0.354	0.36
Model 10	0.364	0.36	0.361
Model 11	0.152	0.352	0.37
Model 12	0.382	0.366	0.375
Stacking model	**0.321**	**0.337**	**0.351**

IX. Conclusion and Future Scope

In this paper, news popularity is predicted using enriched features and ensemble techniques. We have exploited the three categories of features to predict articles' popularity, i.e., Meta features, Text features, and Image features. In the feature extraction phase, various natural language processing techniques are used to fetch out readability, Text Meta, TF-IDF, Sentiment, and dictionary text features. For Image features, a Residual network is used for a deep convolution neural network which gives an enormous amount of image features. These features are further reduced using Principal Component Analysis Technique. The accuracy of popularity prediction is measured with the use of the Root Mean Squared Logarithmic Error (RMSLE) performance measure. In the end, Ensemble XGBoost Model has been used to ensemble the outcomes of all 12 models in a stacking manner which slightly improves the performance of the popularity outcome. Finally, the partial dependence of features on online news popularity is plotted to showcase and understand the insights of feature importance to achieve the research objective.

In the future, work can be extended by incorporating the temporal features for better time series-based performance. Temporal features can be learned using deep learning models- recurrent neural networks or long short-term memory models. Uncertainly, temporal contextual features will help in performance enhancement and can be used for timely fluctuating user preference modeling for numerous social web applications such as recommendation, advertising, and relevant information retrieval. Another extension of work is towards the use of video advertisement-based popularity prediction which requires researchers' attention

References

[1] W. M. Wang, J. Wang, Z. Li, Z. Tian, and E. Tsui, "Multiple affective attribute classification of online customer product reviews: A heuristic deep learning method for supporting kansei engineering," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 33–45, 2019.

[2] J. A. Iglesias, A. Tiemblo, A. Ledezma, and A. Sanchis, "Web news mining in an evolving framework," *Information Fusion*, vol. 28, pp. 90–98, 2016.

[3] E. Stevenson, V. Rodriguez-Fernandez, H. Urrutxua, and D. Camacho, "Benchmarking deep learning approaches for all-vs-all conjunction screening," *Advances in Space Research*, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0273117723000601

[4] J. Del Ser, M. N. Bilbao, I. Laña, K. Muhammad, and D. Camacho, "Efficient fake news detection using bagging ensembles of bidirectional echo state networks," in *2022 International Joint Conference on Neural Networks (IJCNN)*, 2022, pp. 1–7.

[5] J. Huertas-Tato, A. Martín, J. Fierrez, and D. Camacho, "Fusing cnns and statistical indicators to improve image classification," *Information Fusion*, vol. 79, pp. 174–187, 2022.

[6] J. Wang, S. Makowski, A. Cieślik, H. Lv, and Z. Lv, "Fake news in virtual community, virtual society, and metaverse: A survey," *IEEE Transactions on Computational Social Systems*, 2023.

[7] M. Imran, S. Khan, H. Hlavacs, F. A. Khan, and S. Anwar, "Intrusion detection in networks using cuckoo search optimization," *Soft Computing*, vol. 26, no. 20, pp. 10 651–10 663, 2022.

[8] Y. Wang, W. Wang, Q. Chen, K. Huang, A. Nguyen, S. De, and A. Hussain, "Fusing external knowledge resources for natural language understanding techniques: A survey," *Information Fusion*, vol. 92, pp. 190–204, 2023.

[9] X. Yu, Y. Zhu, C. Jia, W. Lu, and H. Xu, "The analysis of green advertisement communication strategy based on deep factorization machine deep learning model under supply chain management," *Expert Systems*, p. e13258.

[10] K. R. Christy, R. Z. Mi, R. Tao, and L. Lu, "Disruptive versus nondisruptive advertising in online streaming video services: How does advertisement placement affect consumer perceptions and ad effectiveness?" *Journal of Advertising Research*, 2023.

[11] Z. Lv, C. Cheng, A. Guerrieri, and G. Fortino, "Behavioral modeling and prediction in social perception and computing: A survey," *IEEE Transactions on Computational Social Systems*, 2022.

[12] G. Zhang, J. Wu, G. Jeon, Y. Chen, Y. Wang, and M. Tan, "Towards understanding metaverse engagement via social patterns and reward mechanism: A case study of nova empire," *IEEE Transactions on Computational Social Systems*, 2022.

[13] T. Hussain, M. Diyan, M. Gogate, K. Dashtipour, A. Adeel, Y. Tsao, and A. Hussain, "A speech intelligibility enhancement model based on canonical correlation and deep learning for hearing-assistive technologies," *arXiv preprint arXiv:2202.04172*, 2022.

[14] A. Ullah and S. Anwar, "The effective use of information technology and interactive activities to improve learner engagement," *Education Sciences*, vol. 10, no. 12, p. 349, 2020.

[15] M. Rehman, I. T. Javed, K. N. Qureshi, T. Margaria, and G. Jeon, "A cyber secure medical management system by using blockchain," *IEEE Transactions on Computational Social Systems*, 2022.

[16] N. Magaia, P. Ferreira, P. R. Pereira, K. Muhammad, J. Del Ser, and V. H. C. de Albuquerque, "Group'n route: An edge learning-based clustering and efficient routing scheme leveraging social strength for the internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19 589–19 601, 2022.

[17] D. Sisodia and D. S. Sisodia, "A hybrid data-level sampling approach in learning from skewed user-click data for click fraud detection in online advertising," *Expert Systems*, vol. 40, no. 2, p. e13147, 2023.

[18] M. Amin, D. Shehwar, A. Ullah, T. Guarda, T. A. Tanveer, and S. Anwar, "A deep learning system for health care iot and smartphone malware detection," *Neural Computing and Applications*, pp. 1–12, 2020.

[19] C. Ieracitano, N. Mammone, M. Versaci, G. Varone, A.-R. Ali, A. Armentano, G. Calabrese, A. Ferrarelli, L. Turano, C. Tebala *et al.*, "A fuzzy-enhanced deep learning approach for early detection of covid-19 pneumonia from portable chest x-ray images," *Neurocomputing*, 2022.

[20] K. Yao, H. Wang, Y. Li, J. J. Rodrigues, and V. H. C. de Albuquerque, "A group discovery method based on collaborative filtering and knowledge graph for iot scenarios," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 1, pp. 279–290, 2021.

[21] C. Foundation, "Coschedule," https://coschedule.com/headline-power-words/, 2019.

[22] S. Foundation, "Sharethrough," https://headlines.sharethrough.com/, 2019.

[23] S. Singhal and P. Tanwar, "An e-commerce prediction system for product allocation to bridge the gap between cultural analytics and data science," *Expert Systems*, p. e13082, 2022.

[24] I. Safder, Z. Mahmood, R. Sarwar, S.-U. Hassan, F. Zaman, R. M. A. Nawab, F. Bukhari, R. A. Abbasi, S. Alelyani, N. R. Aljohani *et al.*, "Sentiment analysis for urdu online reviews using deep learning models," *Expert Systems*, vol. 38, no. 8, p. e12751, 2021.

[25] A. Gupta and H. P. Gupta, "Yogahelp: Leveraging motion sensors for learning correct execution of yoga with feedback," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 4, pp. 362–371, 2021.

[26] Anon, "Uci machine learning repository: Online news popularity data set," Archive.ics.uci.edu./, 2016.

[27] H. Ren and Q. Yang, "Predicting and evaluating the popularity of online news," *Standford University Machine Learning Report*, 2015.

[28] M. Gerasimovic and U. Bugaric, "Enrollment management model: Artificial neural networks versus logistic regression," *Applied Artificial Intelligence*, vol. 32, no. 2, pp. 153–164, 2018.

[29] F. Namous, A. Rodan, and Y. Javed, "Online news popularity prediction," in *2018 Fifth HCT Information Technology Trends (ITT)*. IEEE, 2018, pp. 180–184.

[30] X. Bi and H. Wang, "An enhanced high-order boltzmann machine for feature engineering," *Engineering Applications of Artificial Intelligence*, vol. 78, pp. 37–52, 2019.

[31] S. Oramas, O. Nieto, F. Barbieri, and X. Serra, "Multi-label music genre classification from audio, text, and images using deep features," *arXiv preprint arXiv:1707.04916*, 2017.

[32] K. U. Birant, "Multi-view rank-based random forest: A new algorithm for prediction in esports," *Expert Systems*, vol. 39, no. 2, p. e12857, 2022.

[33] K. Fernandes, P. Vinagre, and P. Cortez, "A proactive intelligent decision support system for predicting the popularity of online news," in *Portuguese Conference on Artificial Intelligence*. Springer, 2015, pp. 535–546.

[34] S. Bhargava and D. S. Ghosh, "Analysis of feature reduction techniques for online news popularity prediction," *SMART MOVES JOURNAL IJOSCIENCE*, vol. 4, p. 11–17, Oct. 2018.

[35] S. Choudhary, A. S. Sandhu, and T. Pradhan, "Genetic algorithm based correlation enhanced prediction of online news popularity," in *Computational Intelligence in Data Mining*. Springer, 2017, pp. 133–144.

[36] J. Rezaeenour, M. Y. Eili, E. Hadavandi, and M. H. Roozbahani, "Developing a new hybrid intelligent approach for prediction online news popularity," *International Journal of Information Science and Management (IJISM)*, vol. 16, no. 1, 2018.

[37] J.-z. Wang, Z. Yan, L. T. Yang, and B.-x. Huang, "An approach to rank reviews by fusing and mining opinions based on review pertinence," *Information fusion*, vol. 23, pp. 3–15, 2015.

[38] Y. Z, J. X, and P. W, "Fusing and mining opinions for reputation generation," *Information Fusion*, vol. 36, pp. 172–184, 2017.

[39] F. G, B. M, and B. A, "Web video popularity prediction using sentiment and content visual features," *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, pp. 289–292, 06 2016.

[40] Z. Y, W. Z, Z. Y, H. M, Y. C, and Q. J, "Exploring popularity predictability of online videos with fourier transform," *IEEE Access*, vol. 7, pp. 41 823–41 834, 2019.

[41] H. Z, Z. M, Z. X, Y. K, T. C, A. Z, K. ., and A, "Automatic understanding of image and video advertisements," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1705–1715, 2017.

[42] Q. D, Z. J, W. Q, and C. G, "Finding competitive keywords from query logs to enhance search engine advertising," *Information & Management*, vol. 54, pp. 531–543, 2017.

[43] A. Goswami and S. H. Chung, "Image-based popularity prediction," Mar. 10 2015, uS Patent 8,977,629.

[44] W. W and Z. W, "Combining multiple features for image popularity prediction in social media," *Proceedings of the 25th ACM international conference on Multimedia*, pp. 1901–1905, 10 2017.

[45] A. A, S. A, and B. S, "Business competitive analysis using promoted post detection on social media," *Journal of Retailing and Consumer Services*, vol. 54, p. 101941, 2020.

[46] D. Liao, J. Xu, G. Li, W. Huang, W. Liu, and J. Li, "Popularity prediction on online articles with deep fusion of temporal process and content features," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 200–207.

[47] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, "Hierarchical attention networks for document classification," in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 2016, pp. 1480–1489.

[48] D. C. Gkikas, K. Tzafilkou, P. K. Theodoridis, A. Garmpis, and M. C. Gkikas, "How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in facebook," *International Journal of Information Management Data Insights*, vol. 2, no. 1, p. 100067, 2022.

[49] A. S. Al-Subhi, "Metadiscourse in online advertising: Exploring linguistic and visual metadiscourse in social media advertisements," *Journal of Pragmatics*, vol. 187, pp. 24–40, 2022.

[50] A. Tarr, J. Hwang, and K. Imai, "Automated coding of political campaign advertisement videos: An empirical validation study," *Political Analysis*, pp. 1–21, 2022.

[51] G. Varoquaux and V. Cheplygina, "Machine learning for medical imaging: methodological failures and recommendations for the future," *NPJ digital medicine*, vol. 5, no. 1, p. 48, 2022.

[52] A. Kontogianni, E. Alepis, and C. Patsakis, "Promoting smart tourism personalised services via a combination of deep learning techniques," *Expert Systems with Applications*, vol. 187, p. 115964, 2022.

[53] shivam, "Kaggle:cnn architectures : Vgg, resnet, inceptionnet, xceptionnet," https://www.kaggle.com/shivamb/cnn-architectures-vgg-resnet-inception-tl, 2019.

[54] shivam bansal, "Kaggle:text-feature-engineering," https://www.kaggle.com/shivamb/text-feature-engineering, 2019.

[55] K. P, G. S, D. S, S. S, and A. A, "Towards contentdependent social media platform preference analysis," *International Journal of Ambient Computing and Intelligence (IJACI)*, vol. 11, pp. 30–47, 2020.

[56] W. Z, S. C, V. Den, and H. A, "Wider or deeper: Revisiting the resnet model for visual recognition," *Pattern Recognition*, vol. 90, pp. 119–133, 2019.

[57] K. K, "Face recognition using principle component analysis," *International Conference on Computer Vision and Pattern Recognition*, vol. 586, p. 591, 1996.

[58] J. Estrada-Jiménez, J. Parra-Arnau, A. Rodríguez-Hoyos, and J. Forné, "On the regulation of personal data distribution in online advertising platforms," *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 13–29, 2019.

[59] O. Zammit, S. Smith, D. Windridge, and C. De Raffaele, "Reducing the dependency of having prior domain knowledge for effective online information retrieval," *Expert Systems*, p. e13014, 2022.

[60] D. Shi, J. Zurada, W. Karwowski, J. Guan, and E. Çakıt, "Batch and data streaming classification models for detecting adverse events and understanding the influencing factors," *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 72–84, 2019.

[61] Z. K. Zandian and M. R. Keyvanpour, "Feature extraction method based on social network analysis," *Applied Artificial Intelligence*, vol. 33, no. 8, pp. 669–688, 2019.

[62] A. Balali, M. Asadpour, and H. Faili, "A supervised method to predict the popularity of news articles," *Computación y Sistemas*, vol. 21, no. 4, pp. 703–716, 2017.

[63] J. N. Farr, J. J. Jenkins, and D. G. Paterson, "Simplification of flesch reading ease formula." *Journal of applied psychology*, vol. 35, no. 5, p. 333, 1951.

[64] R. Flesch, "Flesch-kincaid readability test," *Retrieved October*, vol. 26, no. 3, p. 2007, 2007.

[65] G. H. Mc Laughlin, "Smog grading-a new readability formula," *Journal of reading*, vol. 12, no. 8, pp. 639–646, 1969.

[66] H. Roberts, D. Zhang, and G. S. Dyer, "The readability of aaos patient education materials: evaluating the progress since 2008," *JBJS*, vol. 98, no. 17, p. e70, 2016.

[67] J. Anderson, "Lix and rix: Variations on a little-known readability index," *Journal of Reading*, vol. 26, no. 6, pp. 490–496, 1983.

[68] I. Chaturvedi, E. Cambria, R. E. Welsch, and F. Herrera, "Distinguishing between facts and opinions for sentiment analysis: Survey and challenges," *Information Fusion*, vol. 44, pp. 65–77, 2018.

[69] B. Wang, A. T. Lin, W. Zhu, P. Yin, A. L. Bertozzi, and S. J. Osher, "Adversarial defense via data dependent activation function and total variation minimization," *arXiv preprint arXiv:1809.08516*, 2018.

[70] G.-H. Kim and S.-H. Kim, "Variable selection for artificial neural networks with applications for stock price prediction," *Applied Artificial Intelligence*, vol. 33, no. 1, pp. 54–67, 2019.

[71] J. F. Sánchez-Rada and C. A. Iglesias, "Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison," *Information Fusion*, vol. 52, pp. 344–356, 2019.