Paper:

Arora, A., Bansal, S., Kandpal, C., Aswani, R. & Dwivedi, Y. (2019). Measuring social media influencer index- insights from facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services, 49*, 86-101.
http://dx.doi.org/10.1016/j.jretconser.2019.03.012

Measuring Social Media Influencer Index- Insights from Facebook, Twitter and Instagram

Anuja Arora[1], Shivam Bansal[2], Chandrashekhar Kandpal[2], Reema Aswani[3], Yogesh K. Dwivedi[4]
[1]Jaypee Institute of Information Technology, Noida
[2] Exzeo Software Private limited
[3]Indian Institute of Technology Delhi, New Delhi
[4]Swansea University, Wales, UK

Abstract
The growth of social media has completely revamped the way people interact, communicate and engage. These platforms play a key role in facilitating greater outreach and influence. This study proposes a mechanism for measuring the influencer index across popular social media platforms including Facebook, Twitter, and Instagram. A set of features that determine the impact on the consumers are modelled using a regression approach. The underlying machine learning algorithms including Ordinary Least Squares (OLS), K-NN Regression (KNN), Support Vector Regression (SVR), and Lasso Regression models are adapted to compute a cumulative score in terms of influencer index. Findings indicate that engagement, outreach, sentiment, and growth play a key role in determining the influencers. Further, the ensemble of the four models resulted in the highest accuracy of 93.7% followed by the KNN regression with 93.6%. The study has implications across various domains of e-commerce, viral marketing, social media marketing and brand management wherein identification of key information propagators is essential. These influencer indices may further be utilized by e-commerce portals and brands for the purpose of social media promotion and engagement for larger outreach.

Keywords: social media analytics; influencer; Regression modelling; internet marketing

1. Introduction

Latest technological advancements such as Internet of Things (IOT) (Taylor, Reilly, & Wren, 2018; Lo, & Campos, 2018), Internet of Everything (IOE) (Zwick & Denegri-Knott, 2018), Mobile applications and Social media (Alalwan et al., 2018; Shareef et. al., 2017; 2018; 2019; Shiau et al. 2017; 2018) have brought number of decision making challenges for digital marketing industries. Specifically, Social media platforms have become essentially a medium not only for communication among individual but also for several aspects of business sectors which includes decision making process (Choi, 2017), knowledge-based decision support systems (Chen et al., 2012; Ibrahim et al., 2016), brand promotions (Kaplan & Haenlein, 2010; Lipsman et al., 2012), brand marketing (Aggrawal et al., 2017a; Kapoor et al., 2018), brand and product co-creation (Kamboj et al. 2018; Rathore et al., 2016), product diffusion (Aggrawal et al., 2017b) etc. In the current scenario where there is a constant race for content promotion and propagation, organizations are leveraging the power of social media for reaching out to the masses (Hanna et al., 2011; Kietzmann et al., 2011). It is known to be beneficial in various domains of business and management including social commerce (Chen et al., 2017), e-governance (Dwivedi et al., 2017; Vakeel&Panigrahi, 2018), political marketing (Grover et al. 2018; Kapoor& Dwivedi, 2015) and digital marketing (Aggrawal et al., 2017; Alalwan et al. 2017; Dwivedi et al., 2015; Pintadoet al., 2017; Parsons&Lepkowska-White, 2018).

Now-a-days, social media content has been used by various brands for competing with the competitors, promoting products and offers, and maintaining a reputation among the stakeholders (Brennan & Croft, 2012; Chen, 2013). However, it often becomes difficult for these brands to actually monitor the impact of the brand positioning moves adopted by them (Klostermann et. al., 2018; Pike et. al., 2018). Henceforth, this high volume data is changing the landscape of digital marketing and raised great challenges to turn this brand marketing data into business insights using analytical modelling and management techniques. Social Influencer (SI) index is one such strategy through which brands can discover the right influencers based on their requirements for their brand promotion (Booth&Matic, 2011; Baldus, 2018). Social Media Influencers are users those have highly established credibility for a specific industry like Bollywood (Hearn et. al., 2016), Telecom (Doyle, 2008), News, etc. These social media influencers have connections with large audience and others can also support and trust them due to their admirable authenticity and position (Lou & Yuan, 2018).

It becomes critically essential for brands to identify the right influencers across the web through social media to promote their products and services (Booth &Matic, 2011; Huang et al., 2014). Brands can directly leverage this to improve and enhance public relations by promoting their offerings for higher engagements (De Vries et al., 2011). Identification of social media influencer can be the most important influence marketing strategies to increase the brand's influence on the target audience via their influencers (Lou & Yuan, 2018). Social media influence thus plays a key role in this context at different levels (Romero et al., 2011; Aggrawal et al., 2017). Businesses need social influence to connect with their existing and prospective customers (Mangold & Faulds, 2009). It is an essential requirement for greater interaction and engagement with the potential customers (De Vries, Gensler, & Leeflang, 2012). Further, it can also be beneficial for increasing the visibility in various online communities subsequently leading to a greater outreach (Yang & Kent, 2014).

Influence is the ability to drive action and receive people's engagement on a post which is shared by a strong social influencer on social media or in real life(Freberg et al., 2011). Since the internet is now flooded with large number of influencers - celebrities, athletes, musicians etc., it is necessary to cut through the noise and identify the right category of influencers at the right time (Gillin, 2008). However, computing the influencer index (Morone et. al., 2016) is not a straightforward task and requires the assessment of many data points captured from various sources. Moreover, social media data is also not structured in nature. Though it is available in plenty it needs the right approach to dissect the data into meaningful features (Kiss & Bichler, 2008). Further, it is important to use the features from social media data and regress them in order to generate the influence index. The elements which play a major role in the influencer index are - total engagement, total reach, total sentiment, and total growth(Aggrawal et al., 2018).

This research work is done to measure influencer index on varying social media portal. Basic statistical measures which are commonly used by researchers are not able to learn the system. Due to the same, various machine learning regression models- OLS, KNN, SVM and Lasso Regression model are applied to measure influence of various celebrities on various social media applications. Basically, consumers' reactions towards the posts are considered as features. The above mentioned varying learning model will measure the influence of celebrities based on their reactions on the celebrities' posts which is known as feature engineering. Therefore, a solution for influencer indexing using feature engineering and linear modelling techniques is proposed in this work which is a generic approach and can be applied to any social media application to identify niche of influencers. Even, An ensemble learning model has also been developed to measure impact of influences from social media data. The objective to propose this model is to get better influencer index accuracy. This learning model can be adopted by brands to discover influencers based on their specific needs and requirements. The article further in detail discusses the data used in the model, feature engineering techniques, regression models, and lastly, showcases the results in the form of influence indices.

The presentation of research work is divided in six sections- First section discusses the need to model influencer from social media content. Literature on the same direction is discussed in section 2 followed by social media usage for influence identification, identified knowledge gaps and research contributions. Third section provides a theoretical basis to the study and focuses on research questions with reference to hypothesis of the study. Research methodology for the study is discussed in section four which consists of four subsections- data procurement, feature engineering, feature normalization, and regression modelling. Finding and interpretation of all models and influence ranking results are detailed in section 5. Further, contribution to existing knowledge and implications are described in section 6 followed by concluding remark in section 7.

2. Literature Review

The exponential increase in the amount of content generated through social media forces the network participants to strive for greater attention and subsequent influence on the information takers (Trusov, Bodapati, & Bucklin, 2010). Literature highlights that influence can be easily predicted by URL clicks amongst other important metrics (Romero, et al., 2011). It is further evident that people can better leverage the power of social media by paying attention to the content outreach along with focusing on extending networks (Lipsman et al., 2012). Existing studies also focus on different forms of social media including blogs and conclude that these elite media outlets gain immense traction and have a subsequent social influence on the information consumers (Meraz, 2009; Berthon et al., 2012). This section is divided in three subsections which includes Social media usage for influence identification;

2.1 Social media usage for influence identification

Social media platforms have led to entirely new ways of interaction, communication and engagement (Hansen et al., 2011). Because of the availability of plethora of social networking and media options, it is not a surprise that marketing professionals are very actively exploring these platforms for influencing their potential consumers (Hanna et al., 2011).Recent study by (Weeks et. al., 2017) claims that opinion leaders can be influential and can persuade their peers about news, movies, politics, etc.on social media. Everyone has an influence on social media which could be predicted using an individual’s attributes and historical activities (Bakshy et al., 2011). Studies in literature explore the influence and propagation of content through Twitter (Aswani et al., 2017a; Bakshy et al., 2011; Cha et., 2010), Facebook (Aswani et al., 2017b; Cavalli et al., 2011;), GitHub (Bana et al., 2018) and other popular platforms. The impact is evident in various domains including healthcare (McNeill and Briggs, 2014), education (Tess, 2013), business (Qualman, 2010), Coding portal (Bana et al., 2018), fashion marketing (Wiedmann et al., 2010).

Apart from organizations, there are multitude of individuals including celebrities, actors, bloggers, politicians who voice their opinion on these platforms and act as influencers for the masses (Cha et al., 2010; Dix et al., 2010;Fraser and Brown, 2002).They share their opinions, views, experiences and even daily routine activities that are known to influence their viewers, fans and followers across the globe. These influencers use a combination of these platforms for content dissemination and larger outreach. Studies in literature highlight frameworks that describe factors like spread ability, propagation, integration and nexus for content popularity (Mills, 2012).Neystadt etal.(2012) identified social influencers based on usage context and influencer type. Freberg et al., 2010) mentioned influences as third party endorser and they divert audience attitude through various social media platforms and Blogging sites.

2.2 Identified knowledge gaps and research contributions

Influencer marketing is often done by brands for building strong relationships with the consumers via influencers, a strategy that is mutually beneficial to everyone (Woodcock , Green , & Starkey, 2011). The preliminary study in this direction is performed by (Lagree et al., 2017) in relation to the new form of online marketing known as influencer marketing. In their work, an empirical analysis is applied on twitter data. Influence marketing is basically to connect online personas with brands based on trust and engagement of target audiences on regular basis (Childers et al., 2018).With the increase in the number of offerings by various brands, consumers often look out for authenticity from the brands they interact with. For introducing familiarity and trust factor, brands often use influencer experiences shared on both social media and traditional media along with their posts and commercials respectively (Lou & Yuan, 2018). This makes the product more relevant and trustworthy to the consumers.

Literature highlights that the emerging influencer community is exercising significant power over brand perceptions (Childers et al., 2018). In Childers et al. work, influence marketing based research tries to find out insights and perception of influence marketing and for this experiment data is collected by interviewing professionals of 19 advertisement agencies. Lagree et al. in 2017 proposed a diffusion model to overcome online influence marketing with persistence (OIMP) problem and worked on real data gathered from twitter. Most recent work in this direction is done by (Mallipeddi et al., 2018). This work has been done in two directions- selection of influencers and scheduling of influencers' ads on real data from twitter. Further, a polynomial time heuristic model is introduced to provide optimal solution.

These works have further been driven by the rampant growth of social media which acts as the major platform for influencer communication and subsequent engagement (Nabi, O'Cass, & Siahtiri, 2019). McCormick has investigated celebrity endorsement and measured influence of a product endorser (i.e. celebrity) in order to match consumers attitudes and purchase intentions (McCormick, 2016). Further, the selection of influencer is also very important when he/she has to be chosen to be affiliated to the brand. Even, consumer purchase intention is influenced by credibility and parasocial interaction of social media application which is tested on Instagram and Youtube by Sokolova et. al. in 2019 (Sokolova & Kefi, 2019). It is thus critical to establish an influencer index across social media platforms to enable selection of influencers by various brands (Byrne et al., 2017).Undoubtedly, technologies need to be developed in order to identify and trace influencer on the basis of their content dissemination on varying social media portals.

Researchers worked on influence tracing based on context and influencer type but influencers' role on varying social media portals is untouched area and needs in-depth exploration (Mittal et. al., 2017). However, to the best of our knowledge there are no studies in literature that have taken into consideration an integration of metrics from various platforms for measuring the influence on the target audience. The role of machine learning approach to resolve influence indexing problem is also not explored in existing literature.

Based on above facts, this study tries to: (a) identify the ever growing impact of varying social media platforms and its impact on the masses; (b) measure relevant attributes from three popular platforms- Facebook, Twitter and Instagram; (c) compute an influencer index for the top celebrities in the Indian movie industry using identified relevant attributes; and (d) predict the influencer index of top celebrities using machine learning approaches.

3. Theoretical basis and Hypotheses Development

To the best of my knowledge, there is no study in the literature which highlights how the social influence can be measured on social media sites in relation to the users' engagements. Therefore, there is a need to introduce a method which can be used to measure social influence of a celebrity based on how intimately they are able to engage users with their post. Various computing approaches such as Evolutionary algorithm (Agarwal & Mehta, 2018), Nature inspired algorithm (Aswani, Ghrera, , Kar, & Chandra, 2017) and machine learning approaches (Joseph, Sultan,Kar, & Ilavarasan, 2018) are used by researchers as learning mechanism on other social media issues. Classification models are also used to learn the tweets for predicting test data (Joseph, Sultan,Kar, & Ilavarasan, 2018).

Henceforth, machine learning models had been used as the theoretical lens for framing the context of learning model in order to resolve social media content based influence measurement issue. Although, five machine learning standard algorithms- Naïve bayes, k-nearest neighbors, decision trees, support vector machines, and logistic regression are used by Ma et. al. to predict popularity of a new hashtag on Twitter (Ma, Sun, & Cong, 2013). The broad ambit of social media attracted computer scientist to provide support to numerous business related activities such as advertisement (Goeldi, 2011), recommendation (Taneja et al., 2019; Taneja et al., 2018), product popularity prediction (Jamali et al., 2009), etc. In order to glean useful business intelligence knowledge from extracted social media information, researchers used both supervised (He, 2013; Kelly, Vandevijvere, Freeman, & Jenkin, 2015) and unsupervised machine learning techniques (Anshary, & Trilaksono, 2016; Pham & Simoiu, 2016). Dai et. al. proposed a decision support model in which researcher presents a Mining Environment for Decision (MinEDec) framework (Dai et al., 2011). It analyzes unstructured data to gain business intelligence for a specific outcome such as rival tracking, Environment change detection, strategic matrix etc for competitive intelligence. Hence, machine learning approaches have been used by researchers in order to solve business intelligence problem with respect to social media but influence indexing is not attempted yet using ML techniques.Various text mining and analysis tools have also been used to facilitate in this regard such as SPSS Clementine text mining tool, Nvivo9, AMOS 18.

Social media influence indexing is considered appropriate to evaluate potential users those are main source of enhancing post influence and influencers can be exposed based on various measures such as- size of their social media audience, page engagement, and page views Gaining post influence is the top most priority of all brand marketers. Higher influence on a specific social media leads to higher engagement and higher visibility of content and helps in increasing the high order of discussion among users. This helps post to get viral in market and also can be used for online advertisement purpose. The following research questions are investigated in this research study:

RQ1: What is the impact of social media engagement, outreach and sentiment (in discussions) on social influence index?

In recent study, Sokolova and Kefi investigated the persuasive cues of fashion and beauty influencers based on Youtube and Instagram content. Study refers that the audience created para-social interaction with the influencer (Sokolova & Kefi, 2019). This study claims that attitude homophily positively affect para-social interactions. On similar ground, impact of audience engagements on social influencers' content is questionable which is tried to resolve in this research work.

Indeed, social network consumer behaviour is an important and essential factor to recognize social media influence. Furthermore, an attempt is presented to conceptualize the impact of social media engagement, outreach and sentiment of a specific discussion on influencer indexing. Some multidimensional factors with respect to a specific topic posts are: people-talking-about, likes-count, followers, engagement, outreach, posting rate, post-sentiment etc. These factors seem useful to conform to social influence/ prestige. The Bollywood celebrities are chosen as application domain from three well known social media platforms- Facebook, Twitter, and Instagram.

Therefore, to investigate RQ1, following hypothesises are framed:

- **H1:** Average likes on Instagram has maximum impact for defining social influence as compared to other factor of twitter and Facebook.

To test the significance of features of all three social media platforms regarding social influence two models of multiple linear regression (MLR) are applied. These MLR models are- Ordinary Least Squares (OLS), support vector machine regression (SVM). These MLR are used to test which features of different social media has maximum impact for defining social influence the average and whether Instagram likes give maximum impact to social influence.

- **H2:** Total engagements garnered by the post of Instagram are more impactful as compared to twitter engagement and Facebook engagement.

Other than OLS and SVM regression, one more multiple linear regression model- Lasso regression is experimented. All three models prove that Instagram total engagements are putting better impact as compared to twitter and Facebook.

- **H3:** Different features have varying significance for varying social media platforms.

To validate the significance and impact of consumers four multiple linear regression models- OLS, SVM regression, KNN-Regression, and Lasso regression are modelled. These models are used to identify the greatest impact features i.e. factor those affect social influence maximal.

RQ2: How social media engagements had been associated to identify top social media influencers?

Social media contains number of factors which influence customer engagements. Media and content type of posts is the most significant effect examined by Farook et. al.. In their work five factors those significantly affect the influence (Farook & Abeysekara, 2016) are revealed. One more study claims that identification of engagement has the significant impact on customer engagement (Prentice , Han, Hua, & Hu, 2019). Hence, this proves that social media is shaping influencers based on their interaction on various platforms. These reactions controls and influences the consumer behaviour. Undoubtedly, Social media has become the integral part to influence the society.

The multiple linear regression (MLR) models are used to identify top influencers. An ensemble model is experimented in stacking order based on accuracy of numerous applied MLR models. Best accuracy is achieved by ensemble model, hence this is used for influencer identification result.

Hence, to validate RQ2, designed hypothesis is –

- **H4: Social media interactions have strong connection in top influencer identification.**

RQ3: Which social media platform contributes more to the influencer index and how much accuracy has been achieved in social influence index assignment?

Social media content popularity comparative studies are performed by numerous researchers. These studies claim the method to identify the most suitable social media platform to a specific kind of post (Kaushal, Chandok, Jain, Dewan, & Kumaraguru, 2017; Sokolova, & Kefi, 2019; Macarthy, 2018). Boss et.al. presented an approach to track and measure influence of social networking members for e-commerce sites (Boss et. al., 2018). Henceforth, a platform is required in order to measure influence on varying social media to measure which social media is preferable for what kind of content. In order to influence consumers, celebrity (brand) ends of posting content across multiple social media platforms. For each social media, a specific celebrity influence can be identified using different set of features. Different social media sites have variation in influence of a specific celebrity. Therefore, a framework is needed to compute influence on varying social media portals to distinguish which platform is having which celebrity as most influential. To dig into the same, celebrities and consumers actions on OSNs should be closely monitored and analyzed. Consumers are exposed to diversity of opinion on varying social media platforms which helps them to refine their thought. Furthermore, this diversity of social media causes the variation in influence on different OSNs. To measure this, a detailed lab experiment for consumer reaction to measure celebrity’s influence to assess effectiveness and accuracy of influence index on varying portals is must.

The elementary functionalities of online social network differ. The major OSNs are - facebook as relationship network, Instagram as media sharing network and Twitter as social publishing network. Celebrities end up posting multiple contents across these platforms while availing these services. Indeed, celebrities post content on multiple OSN based on their popularity. Each and every celebrity has variable influence on varying social media sites but still no approach exist to measure social influence on all social media applications. On different social media platform, influence of each one is measured with a set of weighted attributes in accordance of that specific portal. Based on this hypotheses H5 and H6 are spotted.

- **H5: A celebrity has distinctive exposure across OSNs, thereby contributing different influence index on different OSN.**
- **H6: Diverse user actions on diverse social media play an important role for measuring accurate influence.**

The subsequent section focuses on the analysis to address the identified research questions using a mixed research methodology comprising of social media analytics and machine learning approaches used in the study.

4. Research Methodology

The current study uses a mixed researchmethodology comprising of aspects of social media analytics along with machine learning approaches to compute the influencer index across different social media. Social influence using agent based simulation and regression model has been measured by Chan (Chan, 2017). In this research, agent interaction by exchanging social belief and their aggregated neighbours (social connections) belief is described by linear regression model.

The subsequent sub-sections focus on the details of data procurement, feature engineering, feature normalization, regression modelling and subsequent ranking of influencers. The study uses multiple regression approaches for exploring the best accuracy. Somewhat similar sort of work has been presented bypopescu et al. to explore and measure the impact of student performance using social media engagement. Students' active participation has been explored on three social media tools: wiki, blog, and microblogging (popescu et al., 2016). They have applied multiple linear regression model to predict final grades. Their model also presented that several features have an influence on the grade. Based on this study, the primary reason behind using multiple regression modelling techniques is to identify the variables having greater contribution towards the social influencer index or in other words which features have a positive/higher weight in comparison to others.

4.1 Data Procurement

This sub-section discusses the steps used for data collection and dataset preparation. An end-to-end data procurement pipeline comprising of multiple steps is created for the purpose of procuring data for this study. Every step is linked together and is used to fetch the desired data instances and attributes at regular intervals from variety of sources. Figure 1 illustrates the entire data procurement process through a brief flow diagram for the same.

The procurement process starts with the preparation of input seeds. The seeds serve as the inputs to the different social media platforms considered in the study. As an example, the Twitter handle of the influencers serve as the input seed for data extraction, the user identifier for the influencer is an input seed for Facebook, and finally the Instagram URL can be used for information extraction. To prepare these input seeds for various platforms considered in the current work including Twitter, Facebook and Instagram, several available lists on the web are searched and 1000 seeds for various celebrities (influencers) are manually collated. These seeds are for different celebrities across different categories including international athletes, entertainment actors and others.

The second component of this pipeline is the data connectors which are responsible for pulling out the relevant data from the social media platforms using the defined seed inputs. A Twitter Rest API, Instagram API, and Facebook Graph API are used to fetch the desired data which is saved in as a raw file in JSON format.The social media platforms (Facebook, Twitter, and Instagram) under consideration provideseveral page insights including the type of post that is adequate for engaging people and the number of posts that reach a certain number of users.

Subsequently, the third component focuses on the data parsing layer in which the raw data is parsed to generate relevant metrics used for the analysis. This includes mining the overall followers for the considered influencers, their engagements on posts, the content shared by them and the entire meta-data associated with the post (likes, comments and shares). Lastly, data linking module becomes the final component of the pipeline. Since,thedata for same seed (same personality) is extracted from different platforms, it is essential to collate the information mined from these platforms and bundle them together. This is done by maintaining the seed across all the captured data and using the same as a unique identifier in the dataset.

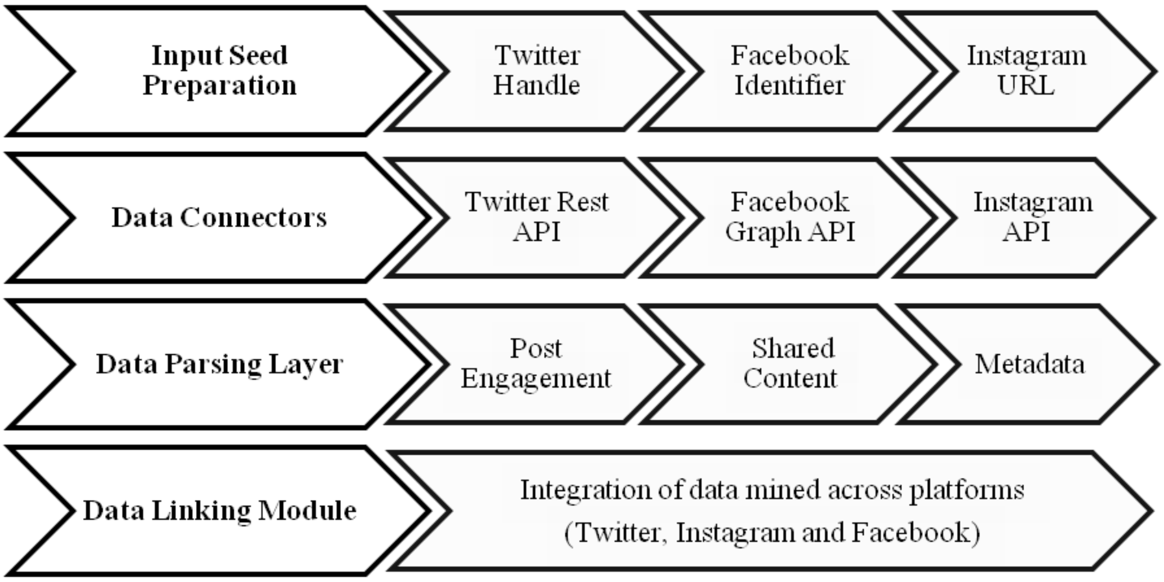

Figure 1: Data Procurement Process Pipeline

Since the entire data collection pipeline is connected together, it has the capability to fetch data at different rates and frequencies. The study mines the data at different frequencies that includes minute-wise, hour-wise and day-wise. The dynamic data pull enabled to create greater number of temporal features associated with the influencer. The complete data comprised of 900 social influencers and their social media attributes obtained from different channels. The model is run on the data which is obtained in a 90 day period window. The summary of data statistics is highlighted in Table 1.

Table 1: Data Statistics Summary

	Twitter API	**Facebook Graph API**	**Instagram API**
Total Seeds	1074	689	783
Total Documents Collected	93,485	37,133	47,113
Data collected minute-wise	234.73	85.05	110.565

Data collected hour-wise	14100.55	5103.25	6634.225

The subsequent section provides a description of the features used to model the social influencer index.

4.2 Feature Engineering

Several features can be directly taken from social media for modelling and analysis. However, since the current study deals with a specific domain/problem statement, a set of desired features have been generated from the existing ones. For the purpose of influencer identification and influencer ranking there are six major components/buckets that are relevant for an influencer's overall rank have been computed as a part of this study. These new features have been created under every social media category. The broad feature categories include the Overall Footprint (OF), Engagements & Outreach (EO), Hourly Engagement Velocity (HEV), Daily Engagement Velocity (DEV), Audience Sentiment (AS) and Posting Rate (PR). These groups and the features categorized under the same are described subsequently.

Overall Footprint (OF):This metric measures the overall presence of an influencer across three channels - Twitter, Facebook and Instagram. This head includes the total Facebook page likes, Instagram page likes, twitter followers count. The raw numbers are bucketed and normalized in a standard range of 0-100 using min-max normalization technique. It also includes data-buckets & ranges created by exploratory data analysis and incorporating domain knowledge.

raw_overall_footprint = sum(twitter followers, Instagram followers, Facebook page likes, Facebook people talking about)

Engagements & Outreach (EO):This metric measures the average engagements per post garnered by an influencer across the three channels. The per post engagements are computed by measuring likes, comments, shares, retweets, and favourites counts on a post created by the influencer. These engagements are aggregated from the influencer post level data of last 30 days from the current date. These engagements are normalized and bucketed in a standard range (using the data-buckets & ranges created specifically for each channel by observing and analysing the complete data of an industry.

Total_outreach = sum(twitter replies, favourites, retweets, likes, comments, shares, reactions)

Hourly Engagement Velocity (HEV):This metric measures the change in engagements per hour from the time of creation of the post. The model considers the average engagements of first hour, second hour, fifth hour and the tenth hour. The final score is computed by aggregating these values, which are further normalized and bucketed in the similar manner as the other metrics.

Daily Engagement Velocity (DEV):This metric is analogous to hourly engagements. However, it measures the change in engagements per day (instead of hour) from the time of creation of the post. A final score is computed by aggregating the average engagements of first, second and seventh day. The values are then normalized using the same procedure.

facebook_span_raw_engagement = sum(likes, comments, shares) * total_span_count / total_posts_collected_every_span

instagram_span_raw_engagement = sum(likes, comments, shares) * total_span_count / total_posts_collected_every_span

twitter_span_raw_engagement = sum(retweets, favourites, replies) * total_span_count / Total_posts_collected_every_span

span -> "hour", "minute", "day", "week"

Audience Sentiment (AS):This metric measures average audience sentiment from audience comments and mentions. Higher the value of this number means more positive sentiment is observed from the audience conversations, lower value of this number implies more negative sentiment is observed from the audience conversations.

The overall positive and negative sentiment in the influencer tweets and posts is computed. The bag of words classifier has been adopted to compute the negative, positive, or neutral sentiment of the data.

Posting Rate (PR): This metric measures the average posting rate for the influencer. It is a measure of the rate at which influencers make the most, if the value is too low means influencer is less active on social media channels, if it is high influencer is most active.

The system is designed to calculate the span dynamically by understanding the overall post distribution of the influencers. For example, if an influencer posts very frequently, his span period is defined as "minute", if it posts moderately slowly, its span period will be "hour", or it posts really slow then its span will be "weekly"

twitter_posting_rate = # of tweets / total_spans
instagram_posting_rate = # of instagram posts / total_spans
facebook_posting_rate = # of facebook posts / total_spans

Table 2 lists the 39 features for every bucket with the category and source of extraction stated against the component being considered. For daily and hourly engagement [plat] may be replaced by fb (Facebook), tw (Twitter) and insta (Instagram) resulting in a set of 18 features on temporal engagement.

Table 2: Feature for Influencer Index Computation

Feature	Category	Source	Acronym	Definition
People Talking About	Overall Footprint (OF)	Page	facebook_ PTA	This is the number of people who have created a story from your Page post in the form of liking, commenting, sharing the page's posts
Total Likes		Page	*_likes	Total Number of Likes on the Facebook page of the brand
Twitter Followers		Timeline	twitter_ followers	Total number of twitter followers of the brand's twitter handle
Instagram Followers		Page	instagram _followers	Total number of followers of the brand page on Instagram
Average Engagement				
Twitter	Engagement (EO)	Tweet	avg_eng_ tw	This number is derived as the average of the sum of retweets and favourites on a tweet over all the tweets made by the

				brand
Instagram		Post	avg_eng_insta	This number is derived as the average of the sum of likes, comments and shares on an Instagram post over all of the Instagram posts of the brand
Facebook		Post	avg_eng_fb	This number is derived as the average of the sum of likes, comments and shares on an Facebook post over all of the Facebook posts of the brand
Aggregated Likes/Favorites/Shares/Comments/Retweets				
LikesFacebook	Outreach (EO)	Post	avg_likes_fb	Average number of likes garnered by the Facebook posts of the brand
Comments Facebook		Post	avg_comments_fb	Average number of comments garnered by the Facebook posts of the brand
Likes Instagram		Post	avg_likes_insta	Average number of likes garnered by the Instagram posts of the brand
CommentsInstagram		Post	avg_comments_insta	Average number of comments garnered by the Facebook posts of the brand
Shares Instagram		Post	avg_shares_insta	Average number of shares garnered by the Instagram posts of the brand
Favourites Facebook		Post	avg_shares_fb	Average number of shares garnered by the Facebook posts of the brand
RT –Twitter		Post	avg_rt	Average number of retweets garnered by the tweets of the brand
Engagement				
Hour1	Hourly Engagement (HEV)	Post	h1_tot_eng_[plat]	Total engagement garnered by the posts on the platform in the 1st hour since it was posted.
Hour5		Post	h5_tot_eng_[plat]	Total engagement garnered by the posts on the platform till the 5th hour since it was posted.
Hour10		Post	h10_till_eng_[plat]	Total engagement garnered by the posts on the platform till the 10th hour since it was posted.
Day1	Daily Engagement (DEV)	Post	d1_eng_[plat]	Total engagement garnered by the posts on the platformtill the 24th hour since it was posted.
Day2		Post	d2_eng_[plat]	Total engagement garnered by the posts on the platformtill two days since it was posted.
Day7		Post	d7_eng_tot_[plat]	Total engagement garnered by the posts on the platformtill first week since it was posted.
Average Post Rate				

Twitter	Posting Rate (PR)	Timeline	avg_post_rate_tw	The rate at which the brand makes tweets. This number is obtained by averaging out the time gap between successive tweets.
Facebook		Page	avg_post_rate_fb	The rate at which the brand makes Facebook posts. This number is obtained by averaging out the time gap between successive posts.
Instagram		Page	avg_post_rate_in	The rate at which the brand makes Instagram posts. This number is obtained by averaging out the time gap between successive posts.
Average Audience Sentiment				
Twitter	Audience Sentiment (AS)	Replies	avg_sent_tw	Average value of overall sentiment computed using the bag of words approach for all the tweets.
Facebook		Comments	avg_sent_fb	Average value of overall sentiment computed using the bag of words approach for all the Facebook posts.
Instagram		Comments	avg_sent_insta	Average value of overall sentiment computed using the bag of words approach for all the Instagram posts.

4.3 Feature Normalization

The collected data and extracted features vary largely in terms of range of values and are on different scales. For instance, the total footprint of an influencer may lie between 100000-500000 while the posting rate may be between 10 and 100. On the other hand the range for audience sentiment may be as low as -2 to 2. In scenarios like the one discussed, if a simple regression metric is used to model the problem, the Audience Sentiment feature will not play any significant role because it is several orders smaller as compared to other features. This feature, which on the contrary, seems insignificant, may actually contain extremely important information which can be useful for computation of the final outcome.Thus, using these features without normalization may bias the outcome in favour of the feature with larger computing the outcome values. If the scales for different features are wildly different, this can have a knock-on effect on the ability of regression models to learn. Hence, in order to make the contribution of these features equal while, it is always a pre-requisite to normalize the data which brings the features on the same scale. A depiction of few instances prior to the normalization process is shown in Table 3.

Table 3: Sample data instances for raw dataset

brand_id	**facebook_PTA**	**facebook_likes**	**twitter_followers**	**instagram_followers**	**max_eng_tw**	**avg_eng_tw**	**max_eng_fb**	**avg_eng_fb**
AmyJackson	21789	2338392	902351	2273038	2911	1007.08	1891	1891
AnoushkaShankar	14298	284841	25644	20831	101	25.60	23424	3282.70
AnushaDandekar	69604	1738888	540137	517705	422	169.12	37182	9396.36

AnushkaSharma	325524	6079571	9146729	8028658	6279	3736.33	84771	19011.57
AyeshaTakia	0	0	389440	153459	0	0	0	0
BipashaBasu	14489	5896143	4955650	3683686	496	235.38	0	0
BrunaAbdullah	1177	155988	64521	213454	28	18.15	1134	352.30
ChitrangdaSingh	0	0	780126	0	1031	569.17	0	0
DeepikaPadukone	424497	32951962	16909445	13751079	5803	3809.38	175798	71343.36

The study uses min-max normalization to scale every feature in the range of 0 to 100. Min-max normalization is often adopted for feature scaling where the values of a numeric range of a feature are reduced to a common scale. Therefore, in order to calculate the normalized value (Z) for an observed value of *x*, Equation 1 is used. The normalized resultssnapshotis depicted in Table 2.

$$Z = \frac{x - min(x)}{[max(x) - min\ (x)]} \tag{1}$$

where*min* and *max* are the minimum and maximum values for the feature *x* given its range.

Table 4: Sample data instances for normalized dataset

facebook_PTA	**facebook_likes**	**twitter_followers**	**instagram_followers**	**max_eng_tw**	**avg_eng_tw**	**max_eng_fb**	**avg_eng_fb**	**max_eng_insta**	**avg_eng_insta**	**avg_likes_fb**
98.06	98.77	98.71	85.43	48.17	84.54	98.2	96.16	48.48	68.48	95.54
70.5	95.26	84	81.22	30.68	50.24	62	49.79	65.8	71.41	48.58
0	0	62.37	64.06	36.33	56.72	0	0	36.11	46.53	0
81.93	80	36.81	54.88	1	1	90.41	95.2	46.25	55.3	95.13
35.95	70.16	68.7	76.92	3	8.75	42.75	40.14	60.14	58.96	40.02
68.51	72.91	67.6	68.73	13.58	35.03	47.44	34.01	55.14	59.92	33.8
67.07	80.76	66.19	81.44	32.13	68.92	67.55	57.91	85.83	95.55	57.65
97.4	97.16	66.01	88.92	21.86	36.73	95.57	88.58	90.24	95.64	88.27
32.1	78.2	29.54	71.67	0	0	0	0	61.44	64.05	0

4.4 Regression Modelling

The initial impetus in this direction is tested by anagnostopoulos et.al., they have applied logistic regression to quantify the extent of social relationship and proves that influence is likely source of correlation with the help of shuffle test (Anagnostopoulos, Kumar, & Mahdian, 2008). A hierarchical classification scheme is proposed in a survey paper which depicts that quantitative assessment methods- influence metrics, information flow and influence model (including machine learning models), network/ graph properties exist in literature to model social influence. Even, qualitative assessment is also possible using social modelling, social matching, and community detection (Razis, Anagnostopoulos, & Zeadally, 2018).

The problem under consideration in this study is surrounding influencer indexing which is a classical regression problem. When it comes to user engagement on social media, some social media users have higher engagement on these platforms and tend to tweet/post more often as compared to others.

Thus, the features used for the prediction of social influence has the values in continuous ranges and so is our target variable, influencer index becomes a continuous variable to estimate. Regression analysis comes as the perfect choice to solve the problem at hand. The concept of regression expresses a statistic connection indicating the average regression on the behaviour of variables. The target variable in this case is computed by combining influencer lists for actual brands. The study uses a collection of influencer data for specific brands across different industries including Entertainment, Sports and Publishing amongst others. The list is collated by combining 43 Indian brands comprising of a combined influencer list of about 1000 Celebrities, Bloggers and YouTubers.

The model is adopted to explain the variation of the influencer index across instances. The variation of the dependant variable is explained by computed its covariance with the independent variables. The independent variables in the current study include the average Instagram likes, Average Tweets, and Facebook posts to name a few. A Multiple Linear Regression (MLR) model is used to compute the influencer index. Equation 2 describes the mathematical model.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots\ldots + \beta_k x_k + \varepsilon \quad (2)$$

where, y is the explained variable; $x_1, x_2 \ldots\ldots x_k$are the$k$ explanatory variables, $\beta_0, \beta_1, \ldots\ldots \beta_k$ are the model parameters and ε is the specification error being the difference between the true and the specified model.

Further, to model the above defined regression problem the study uses three primary implementations including Ordinary Least Squares (OLS) (Craven & Islam, 2011), K-NN Regression (KNN) (Hastie &Tibshirani, 1996), Support Vector Regression (SVR)(Basaket al., 2007) and Lasso with cross validation (Tibshirani, 1996). While adopting such models for regression problems, the main motive is to identify the Best Linear Unbiased Estimator (BLUE). The basic idea is toidentify which variables have a greater impact in creating social influencer index or in other words which features are more important having a positive/higher weight in comparison to others. Thus, in order to achieve the desired results the study attempts to analyse the data by regressing the target variable (influencer index*)* using OLS, KNN, SVR and Lasso.

For the initial comparison among the identified models, Grid searching and Parameter Grid is applied to text the model performance based on data learning and relationship with the target variable. The test takes into consideration the default parameters like C,γ, number of neighbours etc.Grid Searching and Parameter Grid are applied to test the performance of different models by analyzing the data learning and relationship with the target variable. Each model is evaluated on different parameters including accuracy (R-Squared Scores), Mean Absolute Error (MAE), Mean Squared Errors (MSE) and the feature coefficients. The parameter tuning is done using model selection Parameter Grid mechanism havinggrid of parameters that possess discrete number of values for each.The weights for the model coefficients and intercepts are extracted and plotted for linear kernels.The subsequent sub-sections discuss in detail the three regression models adopted in the study.

4.4.1 Ordinary Least Squares (OLS)

The MLR model adopted in the current study, is oftenan ideal choice while modelling the linear relationship between a dependent variable (Target) and one or more independent variables (Predictors) (Andrews, 1974). MLR is based on OLS, the model is fit such that the sum-of-squares of the differences of the observed and predicted values is minimized. The MLR model is based on several assumptions (e.g., errors are normally distributed with zero mean and constant variance). Provided the assumptions are satisfied, the regression estimators are optimal, The optimality is judged by the fact that the estimators areunbiased (expect and true value of the estimator are same), efficient (variance is small as compared to other estimators), and consistent (estimator bias and variance tend to approach zero as the sample size approaches infinity). The square of the determination coefficient (Det_{Ceof})in Equation (3) describes the proportion of variance of the dependent variable explained by the regression model.

$$Det_{Ceof}{}^{2} = \frac{SumSqTot}{SumSqReg} = 1 - \frac{SumSqEr}{SumSqTot} \qquad (3)$$

Where, $SumSqTot, SumSqReg$ and $SumSqEr$are representatives of sum of squares total, regression and errors respectively. The same are defined by Equation (4), (5) and (6).

$$SumSqTot \ = \sum(y - \overline{y})^2 \qquad (4)$$

$$SumSqReg \ = \sum(y' - \overline{y'}\ )^2 \qquad (5)$$

$$SumSqEr \ = \sum(y - y')^2 \qquad (6)$$

For the regression model to be perfect, the $SumSqEr$isideally zero while the $Det_{Ceof}{}^{2}$ is 1. On the contrary, if the regression model is a total failure, $SumSqEr$ and $SumSqTot$ become equal and no variance is explained by the regression making the value of $Det_{Ceof}{}^{2}$zero. However, it is important to keep in mind that there is no direct relationship between high determination coefficientand causation.

4.4.2 Support Vector Regression (SVR)

Support Vector Machine can be applied not only to classification problems but also to the case of regression. It comprises of all the main features that characterize maximum margin algorithm which is a popular nonlinear function that used for linear learning machine mapping into high dimensional kernel induced feature space. The capacity of the system is controlled by parameters that do not depend on the dimensionality of feature space.Similar to how the classification approach works there is motivation to seek and optimize the generalization bounds given for the regression model.

The loss function often referred to as the epsilon intensive function is used since it is known to ignore errors. This is also known to reach a globally optimum solutionand simultaneously ensures a reliable generalization bound. In addition to this, SVR presents the solution using small subset of training points whichprovides enormous computational advantages. The dataset is scaled to train the regression model using linear kernel as expressed in Equation 7.

As a part of tuning the model the grid search functionality is used to test the model’s accuracy and lastly the model with best hyper-parameters in the grid search is adopted for computing the influencer index. Figure 3highlights the features significant features along with the weights.

$$f(x,w) = \ \sum_{j=1}^{k} w_j g_j(x) + \ b \qquad (7)$$

Where, $g_j(x)$ denotes a set of nonlinear transformations, and b is the bias that can be dropped in case of data having zero mean.

4.4.3 K-NN Regression

K-nearest neighbours (KNN) is amongst the popular yet most simple algorithmsthatpredicts the numerical target based on a similarity measure which is often any distance functions. Over the decades KNN has been used in statistical estimation and pattern recognition as a popular and efficient non-parametric technique.A simple implementation of KNN regression is to calculate the average of the numerical target of the 'k' nearest neighbours. Another approach uses an inverse distance weighted average of the nearest neighbours. The regression variant adopts the same distance functions as the KNN classification. The Euclidian (Euc_Dist) and Manhattan (Man_Dist) distance is expressed by Equation 8 and 9 respectively where x and y are the two data instances between which the distance is computed.

$$Euc\_Dist = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (8)$$

$$Man\_Dist = \sum_{i=1}^{k} | x_i - y_i | \qquad (9)$$

4.4.4 Lasso Regression

The Lasso does both parameter shrinkage and variable selection automatically. Since, along with knowing the weights for the features we would also like to know the least important variables which can be eliminated. This will also be eventually informative and thus l_1 regularisation is used. After penalizing (constraining the sum of the absolute values of the estimates), some of the parameter estimates may be exactly zero. The larger the penalty, the further estimates are shrunk towards zero. This is convenient when some automatic feature/variable selection needs to be done or even while dealing with highly correlated predictors, where standard regression will usually have regression coefficients that are exceptionally large.Mathematically, it consists of a linear model trained with $l_1$prior as regularisation model. The objective function ($ObjFunc$) to minimize is expressed in Equation 10.

$$ObjFunc = min \frac{1}{2n_{samples}} || X_w - y ||_2^2 + \alpha||w||_1 \qquad (10)$$

The lasso estimate thus solves the minimization of the least-squares penalty with $\alpha\ ||w||_1$ added, where α is a constant and $||w||_1$ is the l_1-norm of the parameter vector.

5. Findings and Interpretations

This section is divided into three subsections- section 5.1 presents the findings of various MLR techniques to identify the high impact features i.e. rank features in accordance to their significance with social influence. Section 5.2 illustrates the MLR resultant top influencers in terms of their percentile. Section 5.3 shows the comparative influence of a specific celebrity on varying social media platforms and finally section 5.4 presents the accuracy results of various MLR techniques and proposed ensemble technique for social influence indexing.

5.1 High impact features identification Results

Using the MLR techniques (Joseph, Sultan,Kar, & Ilavarasan, 2018), features those are having high association with social influence are identified. We had proposed the following three hypotheses,

- **H1:** Average likes on Instagram has maximum impact for defining social influence as compared to other factor of twitter and facebook.

- **H2:** Total engagements garnered by the post of Instagram are more impactful as compared to twitter.
- **H3:** Different features have varying significance for varying social media platforms.

For testing the hypothesis, Four MLR techniques: OLS, SVM Regression, KNN Regression, lasso regression are applied in order to find out significant feature results. The implementations outcome of all the MLR techniques is shown in further sub sections.

5.1.1. Ordinary least Square (OLS) Results

The OLS experiment utilises this simple yet efficient regression approach for modelling the Social Influence index. The results obtained are after running the regression analysis (Joseph, Sultan,Kar, & Ilavarasan, 2018). It is evident from the statistics summary that the determination coefficient (R-Squared) is 0.894 which is close to 1.

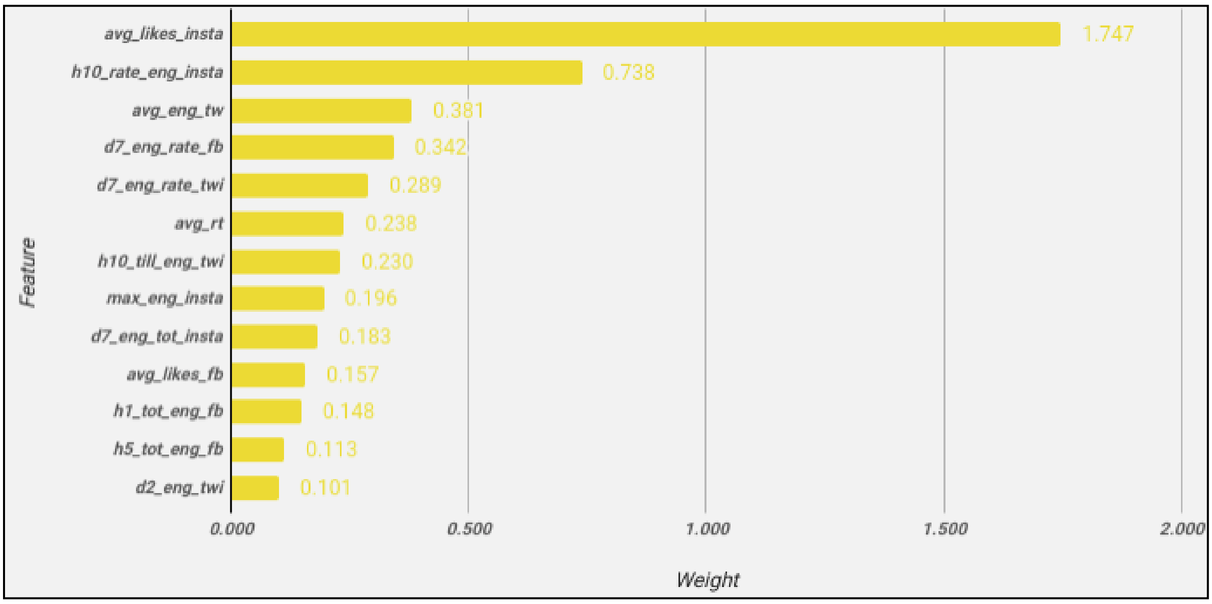

Figure 2: Significant Features obtained by OLS

Further, the model also proves beneficial in identifying the features that have a greater impact on the target social influencer index. Figure 2 is representative of the features that seem to have higher significance on the dependent variable along with the weights using OLS. The average number of likes on Instagram is noticed to have the highest significance, followed by rate of engagement on Instagram with weights 1.747 and 0.738 respectively. It is evident from the graph that average likes on Instagram is the most significant feature (significance value = 1.747) as obtained from the OLS model, followed by engagement on Instagram till the tenth hour since the post was made (significance value= 0.738).This is followed by the average engagement garnered on Twitter (significance value= 0.381) and weekly engagement on Facebook (significance value = 0.342) and Twitter (significance value = 0.289). The average number of retweets on Twitter also shows significant importance. Least significant feature is twitter engagement on day 2(0.101).

5.1.2 SVM Regression Results

SVM Regression results show the features having higher importance on the dependent variable along with weights (Ma, Sun, & Cong, 2013). SVM Regression provided significant features result is depicted in Figure 3.The feature which has highest significance as compared to others is average likes on Instagram (significance value= 1.112). Second highly significant feature is engagement of the post on Instagram till the 10th hour (significance value= 0.748) which is followed by the average engagement on Twitter (significance value= 0.470) and the engagement on Facebook over a week (significance value = 0.365).The results are comparable with OLS model. Even, some features significance value ranking is almost same as given by OLS model.

Ordinary linear regression and SVM regression results support Hypothesis 1 and Hypothesis 2. Hence, H1 and H2 are not rejected and present following as true: Average likes on Instagram has maximum impact for defining social influence as compared to other factor of twitter; Facebook and total engagements garnered by the post of Instagram are more impactful as compared to twitter.

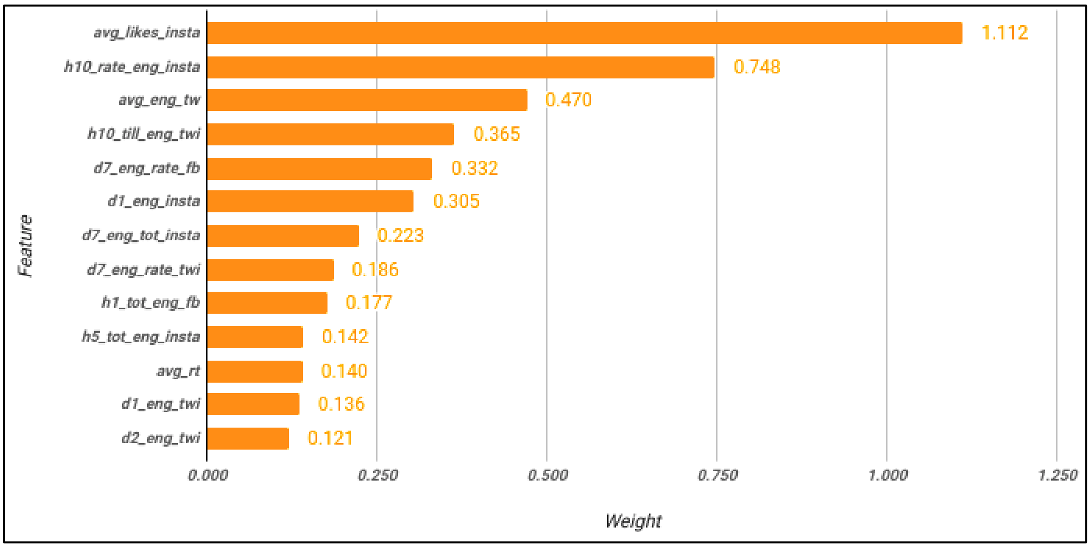

Figure 3: Significant Features obtained by SVR

5.1.3 KNN Regression Results

Choosing the optimal value for 'k' is the most critical aspect of adopting KNN Regression approach. Since, *k* is a critical tuneable hyper-parameter. The model is trained with different values of k and is subsequently checked for accuracy as illustrated in Figure 4 which clearly represents that the accuracy is highest for 2 neighbours.

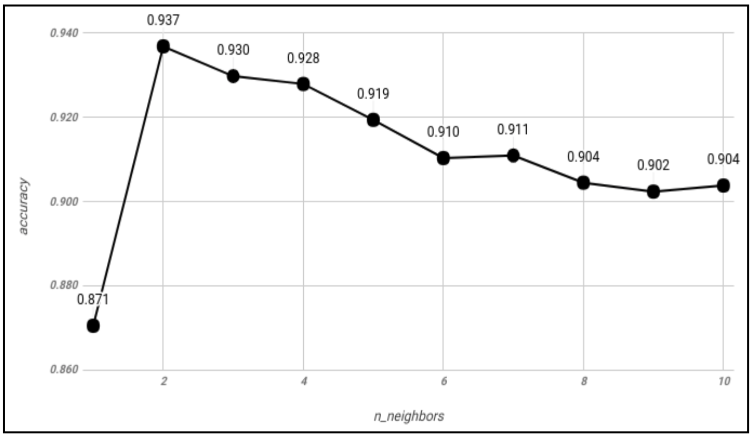

Figure 4: Accuracy for different 'k' neighbours

Further, using KNN with k being 2, the regression model identifies the significant features along with their corresponding weights (Ma, Sun, & Cong, 2013). The features with a greater importance as obtained by the KNN regression model are illustrated in Figure 5.

The total engagement gained by the Facebook post of the influencer up to the 5th hour since it was posted contains maximum significance value = 0.45. further, Engagement garnered on Facebook till the end of the week, Facebook total engagement, engagement over varied intervals on Instagram is at second highest value 0.44.
The results obtained from the KNN approach possess weights having negligible difference varies from 0.45 - 0.36 which reflects the fact that the top ten features are almost equally important while computing the final influencer scores.

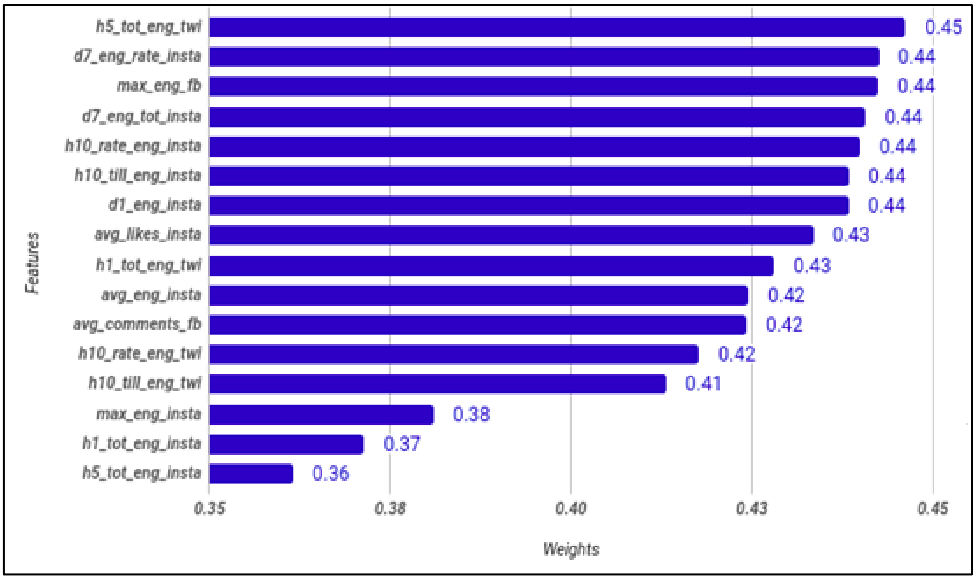

Figure 5: Significant Features obtained by KNN

5.1.4 Lasso Regression Results:
The current study uses the Lasso Regression to train the model and five-fold cross validation for the purpose of result verification. Figure6 illustrates the variation of mean R2 score during cross validation with varying α scores.

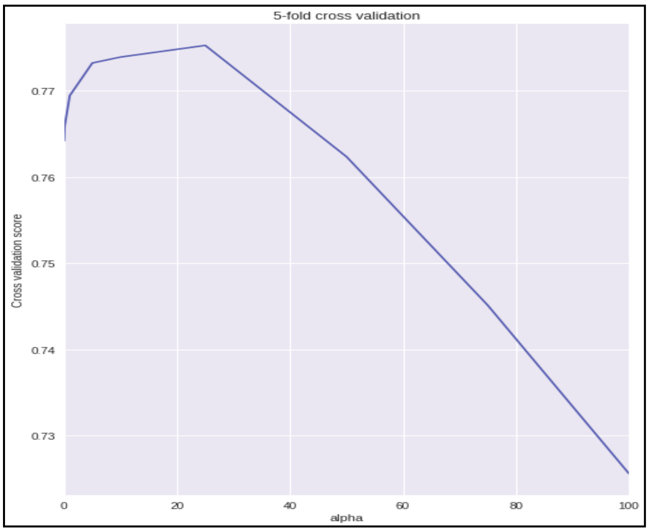

Figure 6: Cross Validation Accuracy vs. α
Figure 7 presents the significant features obtained from Lasso Regression along with their positive and negative weights.

When it comes to the significant features identified by Lasso regression, Result shows that Instagram engagement for a week is the most significant feature with lasso model coefficient value 0.29. Second most significant feature is average twitter engagement (significance value = 0.25). Further, maximum Instagram engagement, Facebook engagement for a week, etc are in decreasing ranking order as shown in figure 7.

The graph also presents negative coefficients wherein highest negative coefficient is average engagement over Instagram (significance value = -0.36). This clearly indicates that the Instagram post is tending to gain higher engagement over time. The average tweets and Instagram followers also showcase negative weights indicating a negative impact while computing the influencer score.

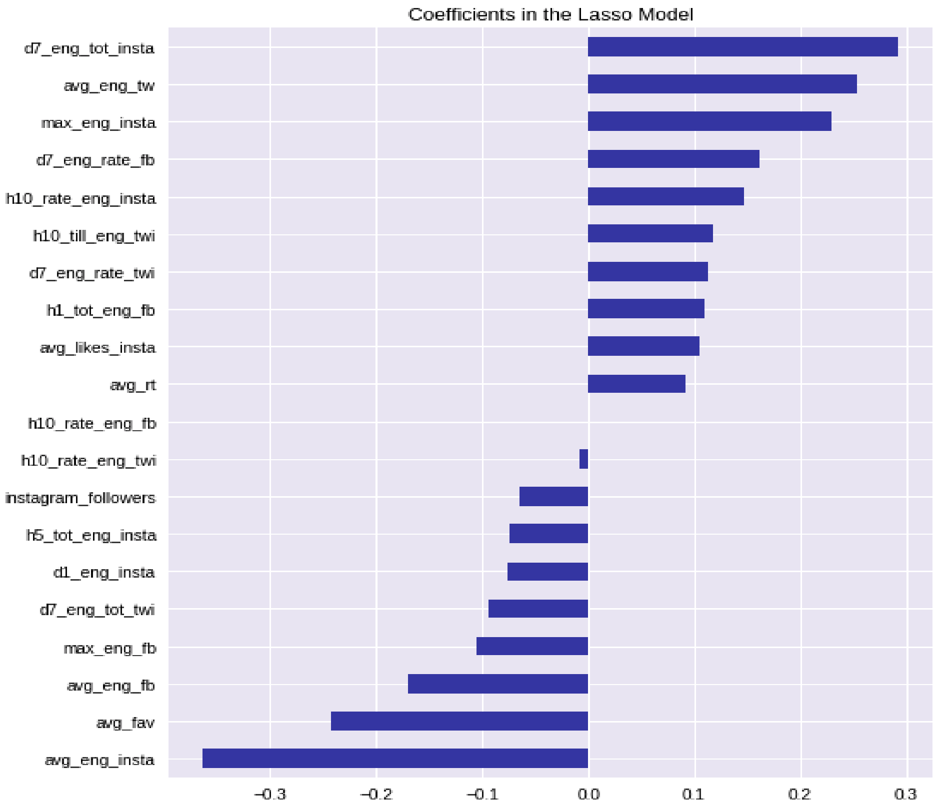

Figure 7: Significant Features obtained by Lasso

Therefore, H3 hypothesis is not rejected. KNN-Regression result and Lasso Regression result shows that features have varying significance value for varying social media platforms which was out hypothesis 3.

5.2 Ranking Results:

In previous section, weights are derived by analysing feature importance of each variable based on the domain knowledge. These are further fined tuned by iterating output results insights. Thus, we hypothesize that social media interaction (i.e. engagement, outreach, etc)have the ability to measure social influence index on social media portals and initialize the process for top influence identification.

- **H4:** Social media interactions have strong connection in top influencer identification.

The weights are validated using linear regression model and ensemble gradient boosting model. For this, Klout score is considerable dependent variable.Klout is a social media analytics website which

rate users in between 1 to 100 based on their online social influence. The study also uses percentiles and z-score to compute the right buckets for a list.
The influencer scores forIndiancelebrities are estimated using the weights of the features generated from the regression models and a list of top 20influencers is computed in terms of their percentile. The result of top 21 influencers are presented in Table 5 Which shows highest rank celebrity as 100 percentile and all influences are computed based on influence percentage.
Apart from ranking the influencers based on percentile, the study also identifies relevant factors comprising of outreach, footprint and sentiment that play a critical role in identifying prominent influencers (section 5.1).

Table 5: Top 20 Influencer Percentile

SalmanKhan	100.00	RiteishDeshmukh	86.25
AkshayKumar	97.58	VarunDhawan	86.13
DeepikaPadukone	94.81	SunnyLeone	84.16
HritikRoshan	93.93	Ranveer Singh	84.00
SonakshiSinha	92.67	SidharthMalhotra	83.74
ParineetiChopra	92.58	DishaPatani	83.58
ShraddhaKapoor	92.01	ShrutiHaasan	82.63
JacquelineFernandez	91.79	KajalAggarwal	81.12
ShahidKapoor	89.43	ArjunKapoor	80.35
PriyankaChopra	88.94	SonamKapoor	80.28
AjayDevgn	88.55		

5.3Comparative Analysis of Influence Index on Distinct Social Media

The elementary functionalities of online social network differ. Few of them the major OSNs are - facebook as relationship network, Instagram as media sharing network and Twitter as social publishing network. Celebrities end up posting multiple contents across these platforms while availing these services. On different social media platform, influence of each one is measured with a different set of attributes in accordance of that specific portal.Based on this hypothesis had been proposed:

H5: A celebrity has distinctive exposure across OSNs, thereby contributing different influence index on different OSN.
Figure 8 illustrates the result of influencer scores for some celebrities across the three social media platforms namely Facebook, Twitter and Instagram. The scores have been computed based on the Outreach, Footprint, Engagement and Sentiment of the influencer. It is evident from the table that DeepikaPadukone has a very high influencer score on Facebook as well as on Twitter while it is comparatively low on Instagram as compared to the other celebrities. On the other side, Anushka Sharma's Twitter influence and Ayesha Takia'sInstagram influence has been low.

brand_id	Facebook	Twitter	Instagram
AliaBhatt	33.38	98.75	76.63
AmritaArora	33.38	68.50	98.75
AmritaRao	39.07	34.88	68.50
AmyJackson	46.84	91.84	34.88
AnoushkaShankar	52.80	38.56	91.84
AnushaDandekar	47.91	82.69	38.56
AnushkaSharma	74.81	26.07	82.69
AyeshaTakia	33.38	55.86	26.07
BipashaBasu	33.38	97.49	55.86
BrunaAbdullah	31.75	47.02	97.49
ChitrangdaSingh	33.38	34.21	47.02
DeepikaPadukone	90.25	98.32	34.21

	Very High
	High
	Moderate
	Low
	Very Low

Figure 8: Platform-wise Influencer Score

5.4 Social Influence Indexing Accuracy Results

The critical influence index of Indian celebrities is obtained from ML regression models. In order to compute Social Influence Index on varying OSN based on different features, four fundamental regression models- Ordinary least square (OLS), Support vector regression (SVR), K-Nearest neighbour regression (KNN-R) and Lasso regression is applied. Even to further improve the accuracy result ensemble of these four basic regression models in stacking manner is validated. To validate this, we had proposed the hypothesis:

- **H6:** Diverse user actions on diverse social media play an important role for measuring accurate influence.

The results for the adopted approaches are compared and evaluated on the basis ofMean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Table 6 reports the details for the same. The KNN approach with 2 neighbours has the lowest MAE of 3.67 while the ensemble model reflects the lowest MSE and RMSE, 32.50 and 5.70 respectively.

Table 6: Regression Metrics

Model	MAE	MSE	RMSE
Ordinary Least Square*(OLS)*	5.72	77.97	8.83
Support Vector Regression *(SVR)*	4.92	80.75	8.98
K-Nearest Neighbours (*k=2*)	3.67	37.56	6.12
Lasso (Alpha=0.1)	5.92	55.8	7.47
Ensemble Model (*OLS, SVR,KNN, Lasso*)	4.49	32.50	5.70

Lastly, accuracy of all MLR model is computed. Accuracy refers to the number of influence index correctly measured to either same or different celebrities from among the total celebrities under study. Figure 9 depicts the comparison plots in terms of accuracy for the models.
It is evident from the graph that ensemble model results in the highest accuracy of 93.7% followed by KNN, Lasso, Linear Regression (OLS) and SVR with accuracies 93.6%, 88.6%, 86.8% and 86.4% respectively. The ensemble and KNN model outperform the remaining approaches in terms of error and accuracy resulting in successful prediction of the social influencer score. These approaches better identify the significant features as compared to the OLS, Lasso and SVR regression models.

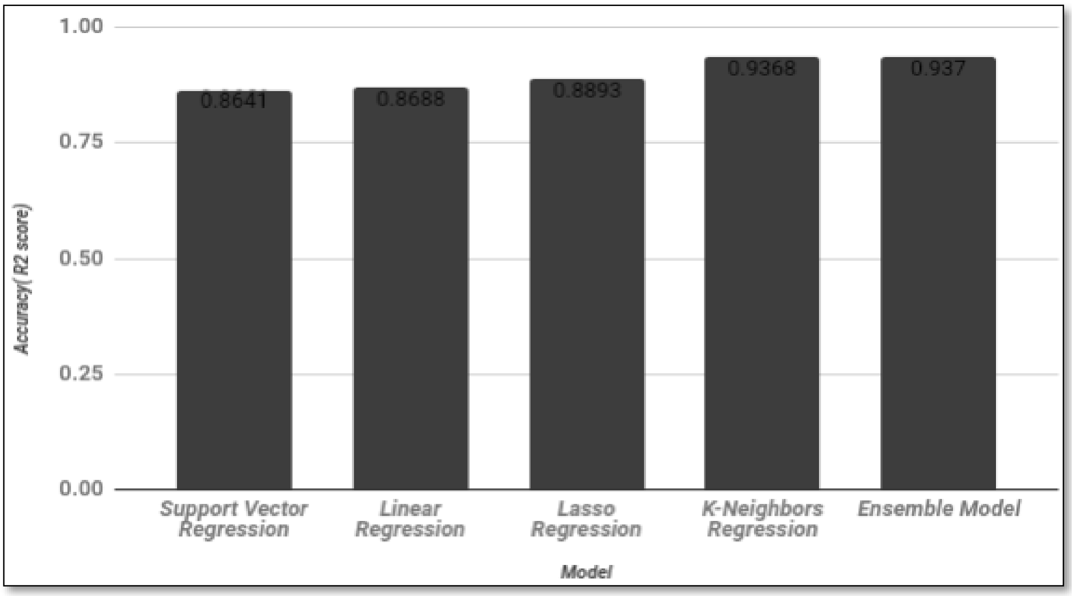

Figure 9: Accuracy of the Models

6. Discussion

Existing studies in literature focus on different forms of social media including blogs and conclude that these elite media outlets gain immense traction and have a subsequent social influence on the information consumers (Meraz, 2009; Berthon et al., 2012; McCormick, 2016; Prentice et; al., 2019). The studies further highlight instances of social media influence and subsequently generated capital (Freberg et al., 2011). However, to the best of our knowledge none of the studies discuss the factors that help in identification of potential influencers. The current study uses a mixed research methodology comprising of both social media analytics and regression analysis to identify significant factors that contribute to social media influencer for selected celebrities (Shareef et. al., 2019). We provided a comprehensive evaluation of contextual features of all three targeted social media platforms. Some of these features are utilized in other studies as well. For example: sentiment, mentions, and hashtags for twitter as features are used in (Ma, Sun, & Cong, 2013); Number of images, Hashtags count, number of filters used , image content length for Instagram are used in (Mittal et. al., 2017). In brief, our evaluation results showed that different features have different weights according to influencer indexing, Influencer indexing computed based on features to influencers/ celebrities and varying influencer/ celebrities have varying influencer index on different social media applications.

The study models and ranks top influencers across three major platforms including Facebook, Twitter and Instagram. This is done based on Outreach, Footprint, Engagement and Sentiment of the influencer. These constructs are computed using attributes mined from the social media profiles of these influencers. Further, none of the studies explore the impact of celebrities' frequency and/or diversity of social media use on their influencer index. The regression model used in the current study tries to identify whether these attributes play a significant role when it comes to computing the social media influencer index. It is observed that the overall influencer score does depend on the individual scores of the different platforms on which the influencers engage. Further, findings also indicate that the frequency of social media usage also increases the social media index in most of the cases.

In addition to this, existing studies explore the impact of user sentiment on firm's equity (Yu et al., 2013). The same can be explained here for an influencer's index. The impact of audience reaction in the form of comments and replies is critical for the influencers. The study explores the effect of overall sentiment scores for computing the influencer index. Lastly, the study also tried to evaluate which social media platform is more dominant when it comes calculating the influencer index. There were not enough instances to establish generalizability of weights that can be attributed with each social media platform. This could be one important area to explore in the future as the current study could not provide sufficient evidences for validating the research question.

6.1 Contribution to Existing Knowledge

The growth of Web 3.0 has enhanced means of interaction, communication and engagement among individuals. In the constant quest for gaining higher than usual traction, social media platforms play a key role. The diffusion of information determines which piece of information cuts through the noise and stands out influencing a larger audience. The current study proposes a scoring mechanism for influencers. The contribution of the study is two-fold. Firstly, the study proposes multiple attributes from social media, categorised into 6 constructs namely Overall Footprint (OF), Engagements & Outreach (EO), Hourly Engagement Velocity (HEV), Daily Engagement Velocity (DEV), Audience Sentiment (AS) and Posting Rate (PR). Further, the study uses regression modelling to compute influencer scores using the identified constructs. These constructs provide a holistic view of different aspects including outreach and engagement which can be useful in multiple use cases. The existing studies can literature can adopt these constructs for understanding social influencers across platforms. The current study through the use of social media analytics

6.2 Implications for Practice

Our study does not directly provide insights for practice but based on correlations, it provides use cases where influencer indices may be useful. Influencer indices can be used by Brand companies or brand marketers and celebrities/ consumers those are intensively associated with social media for promotion purpose. This section is further divided in two subsections which are as follows:

6.2.1 Brand Companies/ Brand Marketers

With the rampant growth of social media usage, brands have started utilising these platforms for enhancing customer engagement and reaching out to a larger audience base. Literature highlights evidence of the impact of social media use in domains like influencer marketing and brand management. These platforms have become increasingly popular forfacilitating engagement, collaborations and drastically impact a brand's reputation (Kim &Ko, 2010; Kim &Ko, 2012). With the plethora of products and services available in the market by different brands, the choice that the consumers have also increases (De Vries et al., 2012). Brands are striving for presence and want to establish themselves for greater outreach and thus invest heavily in influencer marketing. It is very important for these brands to identify influencers that could market their products/services. This makes the current study critically essential for these brands while selecting influencers for their portfolio. The research work for weighted feature finding and influence indices reveals-

- Detailed insights on social media engagement features and which feature is significant on the final social influence indies assignment;
- Investigate the possibility to obtain an interpretable celebrity that makes a specific post highly impactful on a specific social media platform out of Facebook, Instagram, and Twitter.
- Multiple MLR models- OLS, SVM Regression, KNN Regression, and lasso regression is used to build final influence indices model. Based on experiments this is made clear to marketers that out based standard linear regression models- KNN regression provides the highest accuracy for social influence indexing.
- Finally, based on these models, an ensemble model is introduced for marketers. This model is able to provide the highest social influence index generation accuracy 93.7%.

- Marketers are able to identify which social media platform is best for promotion in case of a specifically selected celebrity.

Further, literature highlights several evidences of research being conducted on the success of marketing activities on social media, little is known about which platform is best suited for influencer marketing should there be restriction or limitation in the paid marketing budget.

6.2.1 Celebrities / Consumers:

This study basically investigates the relationship between celebrities' engagement on three popularly known social media platforms. Many celebrity cohort is included in this work for modelling purpose. Undoubtedly, celebrities on social media hold influence over millions of fans. With all this in mind, we suggest few use cases those seems useful for celebrity perspective-

- Social media influence can help celebrity to maintain their status among public.
- Uses social media to bring attention of fans towards their social work and often by pointing societal issues.
- Celebrities may get ideas about which social media platform can viral their news maximally.
- Users/ celebrities may inspire their followers to embrace their personality.
- Celebrities may endorse a specific brand based on their social influence on a specific platform and can earn well by promoting brand on social media.
- Individual can help society in order to make social changes by using their social influence.

The current study provides influencer scores for different platforms and based on a brand's need an influencers score on a particular platform, the best suited choice could be made.

7. Conclusion and Future Research Directions

The current study proposes a mechanism for measuring influencer index across popular social media platforms including Facebook, Twitter and Instagram. The study presents several research questions and in light of those tries to compute a social influencer index. Further, a set of 39 features that help determine the impact on the consumers is modelled using a regression approach. These features have been created under every social media category and are categorized into various sub-heads including Overall Footprint (OF), Engagements & Outreach (EO), Hourly Engagement Velocity (HEV), Daily Engagement Velocity (DEV), Audience Sentiment (AS) and Posting Rate (PR).The features are subsequently analysed using the regression models OLS, KNN, SVR, Lasso Regression and subsequently an ensemble model are adopted to compute a cumulative score in terms of influencer index. The results and findings are indicative of the fact that engagement, outreach, sentiment, and growth play a key role in determining the influencers. The ensemble model outperforms the remaining approaches in terms of error rate and accuracy. The KNN regression also reflects significantly high accuracy almost equal to the ensemble. Further, the study has implications across various domains of e-commerce, viral marketing (Petrescu & Korgaonkar, 2011), social media marketing (Akar & Topçu, 2011), and brand management (Baldus, 2018) where in identification of key information propagators is essential.

The current study does not identify the relative importance of different social media platforms while computing the influencer index. Each platform is given equal weight age while modelling the constructs. This could be one of the future research directions where apart from the constructs a weighted model for platform isolation is used to compute the score. Further, optimization techniques could be used to compute these scores. Evolutionary intelligence including swarm intelligence and bio-inspired computing approaches could be incorporated for finding the optimal value of these weights (Kar, 2016; Aswani et al., 2018a).

Further, Future studies can integrate network metrics like centrality, reciprocity, in-degree and out-degree to better understand the influence of the person on their network. These network related attributes can provide useful insights in terms of information propagation to the social network of influencers on various platforms (Aswani et al., 2018b). Also, a mapping with a personality framework like Big Five could be done to identify personality types of the influencers. The influencers based on their social media activities could be grouped in either of the personality types including extroversion, neuroticism, and openness to experience etc. This could be beneficial in providing a generalised personality of influencers with higher or lower influencer indices enhancing the model adaptability in related domains and use cases.

References

Agarwal, S., & Mehta, S. (2018, August). Social Influence Maximization Using Genetic Algorithm with Dynamic Probabilities. In 2018 Eleventh International Conference on Contemporary Computing (IC3) (pp. 1-6). IEEE.

Aggrawal, N., Ahluwalia, A., Khurana, P., &Arora, A. (2017a). Brand analysis framework for online marketing: ranking web pages and analyzing popularity of brands on social media. *Social Network Analysis and Mining*, *7*(1), 21.

Aggrawal, N., Arora, A., Jain, A., &Rathor, D. (2017b). Product Diffusion Pattern Analysis Model Based on User's Review of E-Commerce Application. In *Hybrid Intelligence for Social Networks* (pp. 227-248). Springer, Cham.

Aggrawal, N., Arora, A., Anand, A., &Irshad, M. S. (2018). View-count based modeling for YouTube videos and weighted criteria–based ranking. In *Advanced Mathematical Techniques in Engineering Sciences* (pp. 149-160). CRC Press

Alalwan, A. A., Rana, N. P., Dwivedi, Y. K., &Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, 34(7), 1177-1190.

Akar, E., &Topçu, B. (2011). An examination of the factors influencing consumers' attitudes toward social media marketing. *Journal of Internet Commerce*, *10*(1), 35-67.

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008, August). Influence and correlation in social networks. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 7-15). ACM.

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*, *16*(4), 523-531.

Anshary, M. A. K., & Trilaksono, B. R. (2016). Tweet-based Target Market Classification Using Ensemble Method. Journal of ICT Research and Applications, 10(2), 123-139.

Aswani, R., Ghrera, S. P., Kar, A. K., & Chandra, S. (2017a). Identifying buzz in social media: a hybrid approach using artificial bee colony and k-nearest neighbors for outlier detection. *Social Network Analysis and Mining*, *7*(1), 38.

Aswani, R., Kar, A. K., & Ilavarasan, P. V. (2018a). Detection of spammers in twitter marketing: A hybrid approach using social media analytics and bio inspired computing. *Information Systems Frontiers*, *20*(3), 515-530.

Aswani, R., Kar, A. K., Aggarwal, S., & Ilavarsan, P. V. (2017b, November). Exploring Content Virality in Facebook: A Semantic Based Approach.In *Conference on e-Business, e-Services and e-Society* (pp. 209-220). Springer, Cham.

Aswani, R., Kar, A. K., Ilavarasan, P. V., & Dwivedi, Y. K. (2018b). Search engine marketing is not all gold: insights from Twitter and SEOClerks. *International Journal of Information Management*, *38*(1), 107-116.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, February). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 65-74).ACM.

Baldus, B. J. (2018). Leveraging Online Communities to Support the Brand and Develop the Community. *Journal of Internet Commerce*, 1-30.

Bana, R., &Arora, A. (2018, August). Influence Indexing of Developers, Repositories, Technologies and Programming Languages on Social Coding Community GitHub. In *2018 Eleventh International Conference on Contemporary Computing* (IC3) (pp. 1-6). IEEE.

Basak, D., Pal, S., &Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews*, *11*(10), 203-224.

Berthon, P. R., Pitt, L. F., Plangger, K., & Shapiro, D. (2012). Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy. *Business Horizons*, *55*(3), 261-271.

Booth, N., &Matic, J. A. (2011). Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: An International Journal*, *16*(3), 184-191.

Boss, G. J., Rick, A. H. I., HUERTAS, L. C. C., & DURAN, E. A. Z. (2018). U.S. Patent No. 9,996,846. Washington, DC: U.S. Patent and Trademark Office.

Brennan, R., & Croft, R. (2012). The use of social media in B2B marketing and branding: An exploratory study. *Journal of Customer Behaviour, 11*(2), 101-115.

Byrne, E., Kearney, J., &MacEvilly, C. (2017).The Role of Influencer Marketing and Social Influencers in Public Health.Proceedings of the Nutrition Society, 76(OCE3).

Cavalli, N., Costa, E., Ferri, P., Mangiatordi, A., Micheli, M., Pozzali, A., ...&Serenelli, F. (2011). Facebook influence on university students' media habits: qualitative results from a field research. In *Media in Transition-unstable platforms: the promise and peril of transition*. US.

Cha, M., Haddadi, H., Benevenuto, F., &Gummadi, P. K. (2010).Measuring user influence in twitter: The million follower fallacy. *Icwsm*, *10*(10-17), 30.

Chen, A., Lu, Y., & Gupta, S. (2017). Enhancing the Decision Quality through Learning from the Social Commerce Components. *Journal of Global Information Management (JGIM)*, *25*(1), 66-91.

Chen, C. P. (2013). Exploring personal branding on YouTube. *Journal of Internet Commerce*, *12*(4), 332-347.

Chen, H., Chiang, R., and Storey V. (2012), Business Intelligence and Analytics: From Big Data to Big Impact, *MIS Quarterly*, 36(4): 1165-1188

Childers, C. C., Lemon, L. L., & Hoy, M. G. (2018). # Sponsored# Ad: Agency Perspective on Influencer Marketing Campaigns. Journal of Current Issues & Research in Advertising, 1-17.

Choi, T. M., Chan, H. K., &Yue, X. (2017).Recent development in big data analytics for business operations and risk management. *IEEE Transactions on Cybernetics*, *47*(1), 81-92.

Craven, B. D., & Islam, S. M. (2011). *Ordinary least-squares regression* (pp. 224-228).Sage Publications.

Dai, Y., Kakkonen, T., & Sutinen, E. (2011). MinEDec: a decision-support model that combines text mining technologies with two competitive intelligence analysis methods. *International Journal of Computer Information Systems and Industrial Management Applications*, 3, 165-173.

De Vries, L., Gensler, S., & Leeflang, P. S. (2012). Popularity of brand posts on brand fan pages: An investigation of the effects of social media marketing. *Journal of Interactive Marketing*, *26*(2), 83-91.

Dix, S., Phau, I., &Pougnet, S. (2010). “Bend it like Beckham”: the influence of sports celebrities on young adult consumers. *Young Consumers*, *11*(1), 36-46.

Doyle, S. (2008). Social network analysis in the Telco sector—Marketing applications. *Journal of Database Marketing & Customer Strategy Management*, 15(2), 130-134.

Dwivedi, Y. K., Rana, N. P., Tajvidi, M., Lal, B., Sahu, G. P., & Gupta, A. (2017, March). Exploring the role of social media in e-government: an analysis of emerging literature. In Proceedings of the 10th International Conference on Theory and Practice of Electronic Governance (pp. 97-106).ACM.

Dwivedi YK, Kapoor KK & Chen H (2015). Social Media Marketing and Advertising. *The Marketing Review*, 15(3), 289-309

Farook, F. S., & Abeysekara, N. (2016). Influence of Social Media Marketing on Customer Engagement. International Journal of Business and Management Invention, 5, 115-125.

Fraser, B. P., & Brown, W. J. (2002). Media, celebrities, and social influence: Identification with Elvis Presley. *Mass Communication & Society*, *5*(2), 183-206.

Freberg, K., Graham, K., McGaughey, K., &Freberg, L. A. (2011). Who are the social media influencers? A study of public perceptions of personality. *Public Relations Review*, *37*(1), 90-92.

Gillin, P. (2008). New media, new influencers and implications for the public relations profession. *Journal of New Communications Research*, *2*(2), 1-10.

Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2018). Polarization and acculturation in US Election 2016 outcomes–Can twitter analytics predict changes in voting preferences. *Technological Forecasting and Social Change*.DoI: https://doi.org/10.1016/j.techfore.2018.09.009.

Goeldi, Andreas (2011)."Website network and adver-tisement analysis using analytic measurement of online social media content." U.S. Patent No. 7,974,983.

Hanna, R., Rohm, A., & Crittenden, V. L. (2011). We’re all connected: The power of the social media ecosystem. *Business Horizons*, *54*(3), 265-273.

Hastie, T., &Tibshirani, R. (1996).Discriminant adaptive nearest neighbor classification and regression.In *Advances in Neural Information Processing Systems* (pp. 409-415).

He, W. (2013a). Examining students’ online interaction in a live video streaming environment using data mining and text mining. Computers in Human Behavior, 29(1), 90-102.

Hearn, A., &Schoenhoff, S. (2016). *From celebrity to influencer* (pp. 194-212). Wiley: London.

Huang, J., Zhang, J., Li, Y., &Lv, Z. (2014). Business Value of Enterprise Micro-blogging: Empirical Study from Weibo. com in Sina. *Journal of Global Information Management*, 22(3), 32-56.

Ibrahim, H., Targio, A., Chang, V., Anuar, N., Adewole, K., Yaqoob, I., Gani, A., Ahmed, E. and Chiroma, H. (2016) The role of big data in smart city. *International Journal of Information Management*, 36 (5), 748-758.

Jamali, S., & Rangwala, H. (2009, November). Digging digg: Comment mining, popularity prediction, and social network analysis. In Web Information Systems and Mining, 2009. WISM 2009. International Conference on (pp. 32-38). IEEE.

Joseph, N., Sultan, A., Kar, A. K., & Ilavarasan, P. V. (2018, October). Machine Learning Approach to Analyze and Predict the Popularity of Tweets with Images. In Conference on e-Business, e-Services and e-Society (pp. 567-576). Springer, Cham.

Kamboj, S, Sarmah, B, Gupta, S & Dwivedi YK (2018). Examining branding co-creation in brand communities on social media: Applying paradigm of Stimulus-Organism-Response. International Journal of Information Management, 39 (April), 169–185.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons*, *53*(1), 59-68.

Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., &Nerur, S. (2018). Advances in social media research: past, present and future. *Information Systems Frontiers*, 20(3), 531-558.

Kapoor, K. K., & Dwivedi, Y. K. (2015). Metamorphosis of Indian electoral campaigns: Modi's social media experiment. *International Journal of Indian Culture and Business Management*, 11(4), 496-516.

Kar, A. K. (2016). Bio inspired computing–A review of algorithms and scope of applications. *Expert Systems with Applications*, *59*, 20-32.

Kaushal, R., Chandok, S., Jain, P., Dewan, P., Gupta, N., & Kumaraguru, P. (2017, September). Nudging nemo: Helping users control linkability across social networks. In International Conference on Social Informatics (pp. 477-490). Springer, Cham.

Kelly, B., Vandevijvere, S., Freeman, B., & Jenkin, G. (2015). New media but same old tricks: food mar-keting to children in the digital age. Current obesity reports, 4(1), 37-45.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011).Social media? Get serious! Understanding the functional building blocks of social media. *Business Horizons*, *54*(3), 241-251.

Kim, A. J., &Ko, E. (2010). Impacts of luxury fashion brand's social media marketing on customer relationship and purchase intention. *Journal of Global Fashion Marketing*, *1*(3), 164-171.

Kim, A. J., &Ko, E. (2012). Do social media marketing activities enhance customer equity? An empirical study of luxury fashion brand. *Journal of Business Research*, *65*(10), 1480-1486.

Kiss, C., & Bichler, M. (2008). Identification of influencers—measuring influence in customer networks. Decision Support Systems, 46(1), 233-253.

Klostermann, J., Plumeyer, A., Böger, D., & Decker, R. (2018). Extracting brand information from social networks: Integrating image, text, and social tagging data. International Journal of Research in Marketing, 35(4), 538-556.

Lagrée, P., Cappé, O., Cautis, B., & Maniu, S. (2017, November). Effective large-scale online influence maximization. In 2017 IEEE International Conference on Data Mining (ICDM) (pp. 937-942). IEEE.

Lipsman, A., Mudd, G., Rich, M., &Bruich, S. (2012). The power of "like": How brands reach (and influence) fans through social-media marketing. *Journal of Advertising Research*, *52*(1), 40-52.

Lo, F. Y., & Campos, N. (2018). Blending internet-of-things (IoT) solutions into relationship marketing strategies. Technological Forecasting and Social Change, 137, 10-18.

Lou, C., & Yuan, S. (2018). Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media. Journal of Interactive Advertising, (just-accepted), 1-45.

Ma, Z., Sun, A., & Cong, G. (2013). On predicting the popularity of newly emerging hashtags in T witter. Journal of the American Society for Information Science and Technology, 64(7), 1399-1410.

Macarthy, A. (2018). 500 Social Media Marketing Tips: Essential Advice, Hints and Strategy for Business Facebook, Twitter, Pinterest, Google+, YouTube, Instagram, LinkedIn, and More!. CreateSpace Independent Publishing Platform.

Mallipeddi, R., Kumar, S., Sriskandarajah, C., & Zhu, Y. (2018). A Framework for Analyzing Influencer Marketing in Social Networks: Selection and Scheduling of Influencers. Fox School of Business Research Paper, (18-042).

Mangold, W. G., &Faulds, D. J. (2009). Social media: The new hybrid element of the promotion mix. *Business Horizons*, 52(4), 357-365.

McCormick, K. (2016). Celebrity endorsements: Influence of a product-endorser match on Millennials attitudes and purchase intentions. Journal of retailing and consumer services, 32, 39-45.

McNeill, A. R., & Briggs, P. (2014, April).Understanding Twitter influence in the health domain: a social-psychological contribution.In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 673-678).ACM.

Meraz, S. (2009). Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication, 14*(3), 682-707.

Mills, A. J. (2012). Virality in social media: the SPIN framework. *Journal of Public Affairs, 12*(2), 162-169.

Mittal, V., Kaul, A., Gupta, S. S., &Arora, A. (2017). Multivariate Features Based Instagram Post Analysis to Enrich User Experience. *Procedia Computer Science*, 122, 138-145.

Morone, F., Min, B., Bo, L., Mari, R., & Makse, H. A. (2016). Collective influence algorithm to find influencers via optimal percolation in massively large social media. Scientific reports, 6, 30062.

Nabi, N., O'Cass, A., & Siahtiri, V. (2019). Status consumption in newly emerging countries: The influence of personality traits and the mediating role of motivation to consume conspicuously. Journal of Retailing and Consumer Services, 46, 173-178.

Neystadt, E. J., Karidi, R., Weisfeild, Y. T., Tennenholtz, M., Radinsky, K., &Varshavsky, R. (2012).U.S. Patent Application No. 13/042,463.

Parsons, A. L., &Lepkowska-White, E. (2018). Social Media Marketing Management: A Conceptual Framework. *Journal of Internet Commerce*, 1-15.

Petrescu, M., &Korgaonkar, P. (2011). Viral advertising: Definitional review and synthesis. *Journal of Internet Commerce*, *10*(3), 208-226.

Pham, T., & Simoiu, C. (2016). Unsupervised Learning For Effective User Engagement on Social Media. arXiv preprint arXiv:1611.03894.

Pike, S., Gentle, J., Kelly, L., & Beatson, A. (2018). Tracking brand positioning for an emerging destination: 2003 to 2015. Tourism and Hospitality Research, 18(3), 286-296.

Pintado, T., Sanchez, J., Carcelén, S., & Alameda, D. (2017). The Effects of Digital Media Advertising Content on Message Acceptance or Rejection: Brand Trust as a Moderating Factor. *Journal of Internet Commerce*, *16*(4), 364-384.

Popescu, P. S., Mihaescu, M. C., Popescu, E., &Mocanu, M. (2016, July). Using ranking and multiple linear regression to explore the impact of social media engagement on student performance. In Advanced Learning Technologies (ICALT), 2016 IEEE 16th International Conference on (pp. 250-254).IEEE.

Prentice, C., Han, X. Y., Hua, L. L., & Hu, L. (2019). The influence of identity-driven customer engagement on purchase intention. Journal of Retailing and Consumer Services, 47, 339-347.

Qualman, E. (2010). Socialnomics: How social media transforms the way we live and do business. *John Wiley & Sons*.

Rathore, A. K., Ilavarasan, P. V., & Dwivedi, Y. K. (2016). Social media content and product co-creation: an emerging paradigm. Journal of Enterprise Information Management, 29(1), 7-18.

Razis, G., Anagnostopoulos, I., & Zeadally, S. (2018). Modeling Influence with Semantics in Social Networks: a Survey. arXiv preprint arXiv:1801.09961.

Romero, D. M., Galuba, W., Asur, S., &Huberman, B. A. (2011, September).Influence and passivity in social media.In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 18-33).Springer, Berlin, Heidelberg.

Shiau W-L, Dwivedi YK & Yang H-S (2017). Co-citation and cluster analyses of extant literature on social networks. International Journal of Information Management, 37(5), 390–399.

Shiau, W-L, Dwivedi, YK & Lai H-H (2018). Examining the core knowledge on Facebook. International Journal of Information Management, 43, 52-63

Shareef, M. A., Mukerji, B., Dwivedi, Y. K., Rana, N. P., & Islam, R. (2019). Social media marketing: Comparative effect of advertisement sources. Journal of Retailing and Consumer Services, 46, 58-69.

Shareef, MA, Mukerji, B, Alryalat, MAA, Wright, A & Dwivedi, YK, (2018). Advertisements on Facebook: Identifying the persuasive elements in the development of positive attitudes in consumers. Journal of Retailing and Consumer Services, 43 (July), 258–268.

Shareef, M. A., Dwivedi, Y. K., Kumar, V., & Kumar, U. (2017). Content design of advertisement for consumer exposure: Mobile marketing through short messaging service. International Journal of Information Management, 37(4), 257-268.

Sokolova, K., & Kefi, H. (2019). Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. Journal of Retailing and Consumer Services.

Taneja, A., & Arora, A. (2019). Modeling user preferences using neural networks and tensor factorization model. International Journal of Information Management, 45, 132-148.

Taneja, A., & Arora, A. (2018). Cross domain recommendation using multidimensional tensor factorization. Expert Systems with Applications, 92, 304-316.

Taylor, M., Reilly, D., & Wren, C. (2018). Internet of things support for marketing activities. Journal of Strategic Marketing, 1-12.

Tess, P. A. (2013). The role of social media in higher education classes (real and virtual)–A literature review. *Computers in Human Behavior*, *29*(5), A60-A68.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.

Trusov, M., Bodapati, A. V., & Bucklin, R. E. (2010). Determining influential users in internet social networks. Journal of Marketing Research, 47(4), 643-658.

Vakeel, K. A., &Panigrahi, P. K. (2018). Social Media Usage in E-Government: Mediating Role of Government Participation. *Journal of Global Information Management*, *26*(1), 1-19.

Weeks, B. E., Ardèvol-Abreu, A., & Gil de Zúñiga, H. (2017).Online influence? Social media use, opinion leadership, and political persuasion. *International Journal of Public Opinion Research*, 29(2), 214-239.

Wiedmann, K. P., Hennigs, N., &Langner, S. (2010). Spreading the word of fashion: Identifying social influencers in fashion marketing.*Journal of Global Fashion Marketing*, 1(3), 142-153.

Woodcock, N., Green, A., & Starkey, M. (2011). Social CRM as a business strategy. Journal of Database Marketing & Customer Strategy Management, 18(1), 50-64.

Yang, A., & Kent, M. (2014). Social media and organizational visibility: A sample of Fortune 500 corporations. *Public Relations Review*, *40*(3), 562-564.

Yu, Y., Duan, W., & Cao, Q. (2013). The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems*, *55*(4), 919-926.

Zwick, D., & Denegri-Knott, J. (2018). Biopolitical Marketing and Technologies of Enclosure (pp. 333-348). The SAGE Handbook of Consumer Culture.