JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY, NOIDA

Department of CSE & IT

Bachelor of Technology, 8th Semester

Determining the influencer index on social media platforms using machine learning algorithms

Group details:

Dhanshree Tejwani 17103194 B4

Soumya Agarwal 17103347 B9

Tanvi Thakur 17103181 B4

Submitted to:

Ankita Wadhwa

Dr. Manish Kumar Thakur

Supervised by:

Dr. Neetu Sardana

Dr. Parmeet Kaur

TABLE OF CONTENTS

PROBLEM STATEMENT

We study the Influence maximisation on social networks and determine the social influencer index of indian celebrities across different professions, on the basis of their presence on social media platforms. We also determine the weightage of various features that have the most effect on the influence maximisation. The social influencer index is determined using various regression models.

IMPORTANCE/RELEVANCE

Consumption patterns and corporate practices have evolved as a result of the internet. The current influencer community is having a massive impact on the marketing strategies of the company's brands. The exponential proliferation of social media platforms through which influencers interact is primarily responsible. Influencer marketing is a strategy of marketing where the emphasis is on a particular group of people or a limited type of person instead of the entire market. It recognizes individuals who have influence over prospective consumers and focuses advertising campaigns on these individuals. Recognizing the ranges of influencers and how they integrate into a brand's social media campaign will ensure improved results and a valuable influence to brand image through social media interaction.

Despite the recent surge of interest in digital influencers as a marketing medium of communication, there is still much to learn about how they can construct a connection with their followers that influence their attitudes and beliefs regarding purchase intention. The Social Influencer (SI) index is one method that enables brands to find the right contributors for their promotional strategy based on customer needs.

RESEARCH WORK SUMMARY

<u>Title</u>: Measuring social media influencer index-insights from Facebook, Twitter, and Instagram

<u>Citation</u>: *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.*

<u>Link</u>: https://www.sciencedirect.com/science/article/abs/pii/S0969698919300128

<u>Summary</u>: People's interactions, communication, and engagement have all changed as a result of the rise of social media. These platforms play a critical role in expanding scope and impact. This research proposes a method for calculating the influencer index on social media sites such as Facebook, Twitter, and Instagram. A regression approach is used to model a collection of features that assess the effect on customers. The underlying machine learning algorithms are modified to calculate a cumulative score in terms of influencer index, including Ordinary Least Squares (OLS), K-NN Regression (KNN), Support Vector Regression (SVR), and Lasso Regression models. Engagement, outreach, sentiment, and development all play a role in deciding the influencers, according to the findings. Furthermore, the KNN regression had the highest accuracy of 93.7 percent, followed by the ensemble of the four models with 93.6 percent. The findings have implications for ecommerce, viral marketing, social media marketing, and brand management, all of which include the identification of key knowledge propagators. These influencer indexes can also be used by e-commerce websites and brands for social media marketing and interaction in order to reach a wider audience.

Title: **Mapping and leveraging influencers in social media to shape corporate brand perceptions**

Citation: *Goodman, Michael B., Norman Booth, and Julie Ann Matic. "Mapping and leveraging influencers in social media to shape corporate brand perceptions." Corporate Communications: An International Journal (2011).*

Link:
https://www.emerald.com/insight/content/doi/10.1108/13563281111156853/full/pdf?casa_token=NAmnLq2A6ssAAAAA:3yD2PZPkuHGukZg-T_Naiguib7OXCzeGmi4XC1Lk2LWMliu6HgtQa44zDrl0GT3UvFYECVrxcLqmkSj3W6XPrlHijjEnB0a8rnAWNGkgOsoj_TtPE8SrGg

Summary: The rapidly expanding social media networks through which influencers interact has given rise to a new influencer culture that wields tremendous control over brand and company perceptions. The "nobodies" of the past have given way to the "somebodies" of today, requiring the attention of communication practitioners seeking constant interaction with targeted customers through the social web's multiple platforms. This paper's aim is to present a method for defining these new "somebodies." This paper examines a customizable valuation algorithm developed to recognise the "new somebodies" who are the influencers driving companies' brand recognition to new heights. The index valuation algorithm takes into account a variety of factors to numerically rank influencers in a social media discussion about a specific business, product, or service. This knowledge aids in the understanding of how these "somebodies" affect conventional target markets, as well as the development of successful outreach campaigns by communications professionals. Integrating the influencer index data into a robust social media strategy for brand equity optimization offers a comprehensive social media approach. The index determines the "conversation points" that should direct engagement with each individual influencer, including topic and tone, and then defines these influencers.

Title: **Influence maximization across heterogeneous interconnected networks based on deep learning.**

Citation: *Keikha, Mohammad Mehdi & Rahgozar, Maseud & Asadpour, Masoud & Faghih Abdollahi, Mohammad. (2019). Influence Maximization across Heterogeneous Interconnected Networks based on Deep Learning. Expert Systems with Applications. 140. 112905. 10.1016/j.eswa.2019.112905.*

Link: https://www.sciencedirect.com/science/article/abs/pii/S0957417419306232

Summary: Because of the rapid growth of online social networks, all of their users are members of several social networks. One of the most important social network research challenges is determining the most active participants. The influence maximisation (IM) challenge seeks to find a small number of users who can transfer the most influence around the underlying network. Most previous studies have focused on a single social network, while people in the real world enter several social networks. As a result, power will propagate through multiple networks by common users. Furthermore, owing to the simplicity of these methods for investigating networks and computing their power diffusion, current works such as simulation-based, proxy-based, and sketch-based approaches suffer from various issues such as scalability, performance, and viability. Furthermore, multiple heuristics are used in the previous algorithms to grab network topology for IM. However, because of their pruning techniques, these approaches lose knowledge during network discovery. A new research direction for researching the IM problem on interconnected networks is proposed in this article. The proposed method learns the feature vectors of network nodes using deep learning techniques while retaining both local and global structural knowledge. Network embedding has not yet been used to solve the IM dilemma, to the best of our knowledge. Indeed, our algorithm employs deep learning techniques for feature engineering in order to retrieve all relevant knowledge for the IM problem in both independent and interconnected networks.

Title: **Setting the future of digital and social media marketing research: Perspectives and research propositions**

Citation: *Dwivedi, Yogesh K., et al. "Setting the future of digital and social media marketing research: Perspectives and research propositions." International Journal of Information Management (2020): 102168.*

Link: https://www.sciencedirect.com/science/article/abs/pii/S0969698919300128

Summary: Consumer behaviour and corporate practises have also evolved as a result of the internet and social media's use. Organizations may benefit from social and digital marketing by lowering prices, increasing brand recognition, and increasing sales. Negative electronic word-of-mouth, as well as disruptive and frustrating online brand presence, pose major challenges. This article puts together the combined wisdom of a number of leading experts on digital and social media marketing issues. The perspectives of the experts include a concise narrative on key aspects of this important subject, as well as perspectives on more relevant issues such as artificial intelligence, augmented reality marketing, digital content management, mobile marketing and advertising, B2B marketing, electronic word of mouth, and ethical issues related to it. This study provides a significant and timely contribution to both researchers and practitioners in the form of challenges and opportunities, in which we illustrate current research weaknesses, outline research gaps, and formulate questions and propositions that can help advance awareness in the field of digital and social marketing.

Title: **Efficient influence maximization in social networks**

Citation: *Chen, Wei, Yajun Wang, and Siyu Yang. "Efficient influence maximization in social networks." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009.*

Link:https://dl.acm.org/doi/pdf/10.1145/1557019.1557047?casa_token=XquW2lqUiXEAAAAA:xvX30aUadSY3aUHjF3DzRo3dAe8y4wJx7fdVotC37Mjfg2QsQ4nYXmwSF34KrEEXgpc_Htam--H4Z8s

Summary: The problem of finding a small subset of nodes (seed nodes) in a social network that can increase the distribution of influence is known as influence maximisation. We look at efficient influence maximisation from two different perspectives in this paper. The first is to develop the original greedy algorithm and make improvements to it in order to reduce its running time even further, and the second is to propose new degree discount heuristics that improve impact distribution. Experiments on two massive academic collaboration graphs collected from an online archival database are used to validate our algorithms. Our results show that (a) our improved greedy algorithm achieves better running time as compared to the improvement of matching influence spread, (b) our degree discount heuristics achieve much better influence spread than classic degree and centrality-based heuristics, and (c) when tuned for a specific influence cascade model, it achieves nearly matching influence thread with the greedy algorithm. Based on our findings, we conclude that fine-tuned heuristics can be able to provide genuinely scalable solutions to the influence maximisation problem, with a pleasing influence spread and a lightning-fast running time. As a result, contrary to the conclusion that conventional heuristics are outperformed by the greedy approximation algorithm, our findings cast new light on heuristic algorithm study.

Title: **Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach**

Citation: *Liu, Shixi, et al. "Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach." Information sciences 306 (2015): 34-52.*

Link: https://www.sciencedirect.com/science/article/abs/pii/S0020025515000729

Summary: Since successful online social network (OSN) influencers can significantly influence consumers' buying decisions through user trust in electronic word-of-mouth (eWOM) marketing, identifying these influencers with respect to user trust relationships has become increasingly relevant. Many recent research, on the other hand, ignore the domain attribute of trust as well as the time-varying existence of social networks, analysing only a static snapshot of a consumer trust network (UTN). This study proposes a research structure that takes into account the dimensions of trust, domain, and time to resolve these issues and investigate this subject in the e-commerce context. Using the time-varying features of multi-type relationships, a time-varying hypergraph is created to model the OSN, and an algorithm is created to extract a domain-aware UTN based on the time-varying hypergraph and user confidence relationships. A novel product review domain-aware (PRDA) approach is conceived, based on the dimensions of trust, domain, and time, that identifies effective influencers and categorize them into three types, namely emerging influencers, holding influencers, and vanishing influencers, based on their popularity status across the life cycle.

<u>Title</u>: **Analyzing the startup ecosystem of India: a Twitter analytics perspective**

<u>Citation</u>: *Singh, Shiwangi, Akshay Chauhan, and Sanjay Dhir. "Analyzing the startup ecosystem of India: a Twitter analytics perspective." Journal of Advances in Management Research (2019).*

<u>Link</u>: https://www.emerald.com/insight/content/doi/10.1108/JAMR-08-2019-0164/full/html

<u>Summary</u>:The purpose of this paper is to use Twitter analytics for analyzing the startup ecosystem of India. The paper examines 53,115 tweets from 15 Indian startups in various industries using descriptive analysis and content analytics techniques from social media analytics. To generate insights for the startup ecosystem in India, the study uses techniques such as the Naive Bayes Algorithm for sentiment analysis and the Latent Dirichlet allocation algorithm for topic modelling of Twitter feeds. The Indian startup ecosystem is leaning toward digital technology, with a focus on people, planet, and benefit, as well as resource availability and knowledge as critical success factors. The study divides tweet emotions into three categories: positive, neutral, and negative. It was discovered that there are more positive than negative feelings in the Indian startup ecosystem. The categorization of the defined keywords into clusters is made possible by topic modelling. In addition, the study concludes that Digital India is the future of the Indian startup ecosystem. The analysis provides a methodology that future researchers can use to extract relevant information from Twitter. Any attempt to use social media research to study India's startup ecosystem is limited. This study attempts to close the gap by examining India's startup ecosystem through the prism of social media outlets such as Twitter.

<u>Title</u>: Influencer marketing: Social media influencers as human brands attaching to followers and yielding positive marketing results by fulfilling needs

<u>Citation</u>: *Cuevas, Leslie M., Sze Man Chong, and Heejin Lim. "Influencer marketing: Social media influencers as human brands attaching to followers and yielding positive marketing results by fulfilling needs." Journal of Retailing and Consumer Services 55 (2020): 102133.*

<u>Link</u>: https://www.sciencedirect.com/science/article/abs/pii/S0969698920300059

<u>Summary</u>: While much of the literature on this subject has focused on the power mechanism that social media influencers (SMIs) use to sway their followers, little is known about their attachment mechanism. Given that social media networks were created to promote personal bonding rather than product or brand recommendations, we hypothesised that social media followers' emotional connection to SMIs sets a precedent that influences their behavioural propensity to embrace the SMIs' endorsements. By focusing on their attachment formation process and its casual influences and efficacy, we were able to bring new attention to the relationship between SMIs and their followers. Study 1 inductively explored the primary causal factors that make followers feel attached to SMIs, both in terms of SMI persona- and content-driven attributes. Study 2 provided empirical evidence by analysing 325 U.S. consumers' responses on how SMIs' personas (i.e., inspiration, enjoyability, and similarity) and content curation abilities (i.e., informativeness) affected followers' perceptions of the SMIs as human brands that fulfil their needs for ideality, relatedness, and competency. This positive emotion fostered by SMIs was transferred to SMIs' endorsements, which in turn positively motivated followers to purchase the products/brands that the SMIs recommended.

Title: **The Impact of The Social Media Influencer Power On Consumer Attitudes Toward The Brand: The Mediating/Moderating Role of Social Media Influencer Source Credibility**

Citation: *Nafees, Lubna, and Christy M. Cook. "The impact of social media power on consumer attitudes toward the brand: The mediating/moderating role of social media source credibility." (2020).*

Link:https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=1409&context=ama_proceedings

Summary: The effect of social media influencer power on customer attitudes toward a brand is discussed in this paper. The study supports the claim that social media influencer power can affect consumer brand attitudes by using naive theories of social influence, consumer socialisation theory, and market signalling theory. The effect of social media influencer power on customer brand attitudes, on the other hand, is thought to be mediated or moderated by the legitimacy of the social media influencer source. As a result, social media influencer source reputation is modelled as being positively linked to the social media influencer's product expertise/competence, goodwill toward the customer, and trustworthiness.

Title: **Social Influence Analysis in Large-scale Networks**

Citation: *Tang, Jie & Sun, J. & Wang, Chi & Yang, Zi. (2009). Social influence analysis in large-scale networks. Social Influence Analysis in Large-scale Networks. 807-816. 807-816. 10.1145/1557019.1557108.*

Link:
https://www.researchgate.net/publication/221653366_Social_influence_analysis_in_large-scale_networks

Summary: Nodes (users, entities) in large social networks are affected by others for a variety of reasons. Colleagues, for example, have a significant impact on one's job, while friends have a significant impact on one's everyday life. The said paper proposed Topical Affinity Propagation (TAP) to model the topic-level social influence on large networks by differentiating the social influences from different angles (topics), quantifying the intensity of those social influences, and estimating the model on real large networks. TAP, in particular, can perform topic-level impact propagation using the effects of any topic modelling and the current network structure. They present several important applications on real data sets using influence analysis, such as 1) what are the representative nodes on a given topic? 2) How do I determine the social effects of nearby nodes on a given node? To put it another way, a user's effect on others is influenced not just by their own topic distribution, but also by the social relationships they have with others. The aim is to create a coherent methodology for social influence analysis that uses both local attributes (topic distribution) and global structure (network information).

Title: **Influence analysis of emotional behaviors and user relationships based on Twitter data**

Citation: *K. Tago and Q. Jin "Tsinghua Science and Technology, vol. 23, no. 1, pp. 104-113, Feb. 2018"*

Link: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8293077

Summary: Users follow and unfollow others for a variety of reasons, including subjects, interests, personality, and first impressions. We consider emotional expression to be one of these factors in this analysis. Tweets are thought to have an impact on a user's relationship with others on Twitter, where emotional expression is common. If a user consistently posts negative messages, his or her followers may become uneasy, which is likely to result in a deterioration of the follow relationship. Take, for example, human interactions in the real world. Since it hurts others' feelings, an offensive term can have a negative impact on human relationships. On Twitter, user relationships are influenced in the same way. In the real world, however, a friendly and bright person has more human relationships than a negative person. Positive Twitter users are more likely to build a deeper friendship. We believe that emotional attitudes would affect user relationships in this analysis. They used a statistical approach to examine the impact of emotional attitudes and user relationships in this analysis. The question here is whether different processes between positive and negative users create the user relationship. This research compares the variations in user relationships between positive and negative users.

Title: **Influence maximization based on activity degree in mobile social networks**

Citation: *Gao, Min & xu, li & Lin, Limei & Huang, Yanze & Zhang, Xinxin. (2020). Influence maximization based on activity degree in mobile social networks. Concurrency and Computation: Practice and Experience. 32. 10.1002/cpe.5677.*

Link:
https://www.researchgate.net/publication/339131838_Influence_maximization_based_on_activity_degree_in_mobile_social_networks

Summary: The problem of influence maximization (IM) has become an important research topic due to the rapid growth of mobile social networks. It attempts to identify a set of nodes, referred to as influencers, contributing to the spread of maximum information. In this article, we present the construction of a social relation graph based on mobile communication data. And we propose a new centrality measure—activity degree to characterize the activity of nodes. By combining the local attributes of nodes and the behavioral characteristics of nodes to measure node activity degree, which can be used to evaluate the influence of users in mobile social networks, we introduce Susceptible-Infected-Susceptible model to simulate the dynamic spreading of information. We take advantage of the two indicators, the degree centrality and the betweenness centrality to get better ranking results. In comparison with spanning graph and initial graph, the results of comparison demonstrate that our algorithm has advantages in the scope of influence propagation.

Title:**The Impact of Social Media Influencers on Purchase Intention and the Mediation Effect of Customer Attitude**

Citation: *Lim, Xin Jean, et al. "The impact of social media influencers on purchase intention and the mediation effect of customer attitude." Asian Journal of Business Research 7.2 (2017): 19-36.*

Link:http://ganj-ie.iust.ac.ir:8081/images/2/22/Ajbr170035.pdf

Summary: The use of social media influencers in advertising is first investigated, especially to generate buzz in younger markets and increase social media coverage in businesses. The aim of this study is to look into the effectiveness of social media influencers, with a focus on source reputation, attractiveness, product match-up, and meaning transfer. It is suggested that consumer attitude mediates both exogenous and endogenous relationships. Purposive sampling was used to collect data, and the dataset of 200 respondents was then analysed using the PLS-SEM methodology. Except for source credibility, all theories are found to be supported. The effects of customer attitude on mediating effects are also calculated. Implications, drawbacks, and recommendations for future study are also addressed in this paper.

Title: **The influence of Twitter on education policy making**

Citation: *Omar, Mwana & Njeru, Alexander & Yi, Sun. (2017). The influence of Twitter on education policy making. 133-136. 10.1109/SNPD.2017.8022712.*

Link:
https://www.researchgate.net/publication/319410031_The_influence_of_Twitter_on_education_policy_making

Summary: In several areas, such as e-commerce, marketing, and e-learning, collecting input for the purpose of evaluating relevant data is critical. However, due to insufficient capacity to manage large amounts of data, governments, businesses, and organisations find it difficult to respond to public input quickly on Twitter because of the vast amount of informal and unstructured data. The study's goals are to: i) use sentiment analysis to analyse the impact of Twitter on education policymaking. (ii) Create a sentiment analysis model that uses the Naive Bayes algorithm to classify the polarity of tweets automatically. (iii) To look at various data visualisation methods that make it easier for education policymakers to consider the sentiment analysis findings. Apart from this, future work can be done to explore and use all the features of twitter.

<u>Title</u>: Analysis of User Network and Correlation for Community Discovery Based on Topic-Aware Similarity and Behavioral Influence

<u>Citation</u>: *Zhou, Xiaokang & Wu, Bo & Jin, Qun. (2017). Analysis of User Network and Correlation for Community Discovery Based on Topic-Aware Similarity and Behavioral Influence. IEEE Transactions on Human-Machine Systems. PP. 1-13. 10.1109/THMS.2017.2725341.*

<u>Link</u>:
https://www.researchgate.net/publication/319361469_Analysis_of_User_Network_and_Correlation_for_Community_Discovery_Based_on_Topic-Aware_Similarity_and_Behavioral_Influence

<u>Summary</u>: The user-generated data, associated with social behaviors, could be viewed as a valuable social resource for constructing dynamic user networking, not only to motivate collaborations among individuals, but also to facilitate information propagation across communities. Compared with most algorithms that consider user networks as static graphs, the key challenges to dynamic social network analysis are as follows: How to track and identify temporal changes and evolving features of social networks in different time frames over time; how to characterize the nature of dynamically changed user networks within a certain social context, when user associations and related information are involved on a large scale; and how to efficiently measure the strength of dynamic networks, which not only represent the hidden structure among various associated linkages, but also reveal value-added rich information in meaningfully structured communities. To address these challenges, a model of dynamic user networking was constructed in this study, and the potential user correlation is analyzed from the user-generated data with social behaviors. In particular, the focus is on identifying two special kinds of social relationships to represent the dynamics among a group of connected users: similarity based and influence based. Both topic-aware similarity and behavioral influence are seamlessly incorporated to refine dynamic user correlations. Specifically, this study focuses on proposing the following. i) A generic user networking model, named dynamically socialized user networking (DSUN) that represents a user's topic-aware features and interactional behaviors in a unified way, to identify and describe implicit and explicit user relationships in a certain social context. ii) A set of measures to analyze, refine, and quantify the social behavior-related user correlations among a group of people, based on topic-aware similarity and behavioral influence. iii) Three algorithms to extract and describe three types of ties, considering influence-based user relationships, which enable the discovery of different kinds of communities that assist in information dissemination and knowledge sharing.

WORK PLAN

1. Understanding
    a. Influence Maximisation
    b. Social Influence Index
2. Data Procurement
    a. Input Seed PreparationPreparation
    b. Data Connectors
    c. Data Parsing
3. Data Engineering and Normalisation
4. Applying Regression models on the dataset obtained
    a. Ordinary Least Squares (OLS)
    b. Support Vector Regression (SVR)
    c. K-NN Regression
    d. Lasso Regression
5. Analysing weights of different features using regression models
6. Comparative Analysis of Influence Index obtained by applying regression models.

WORKFLOW

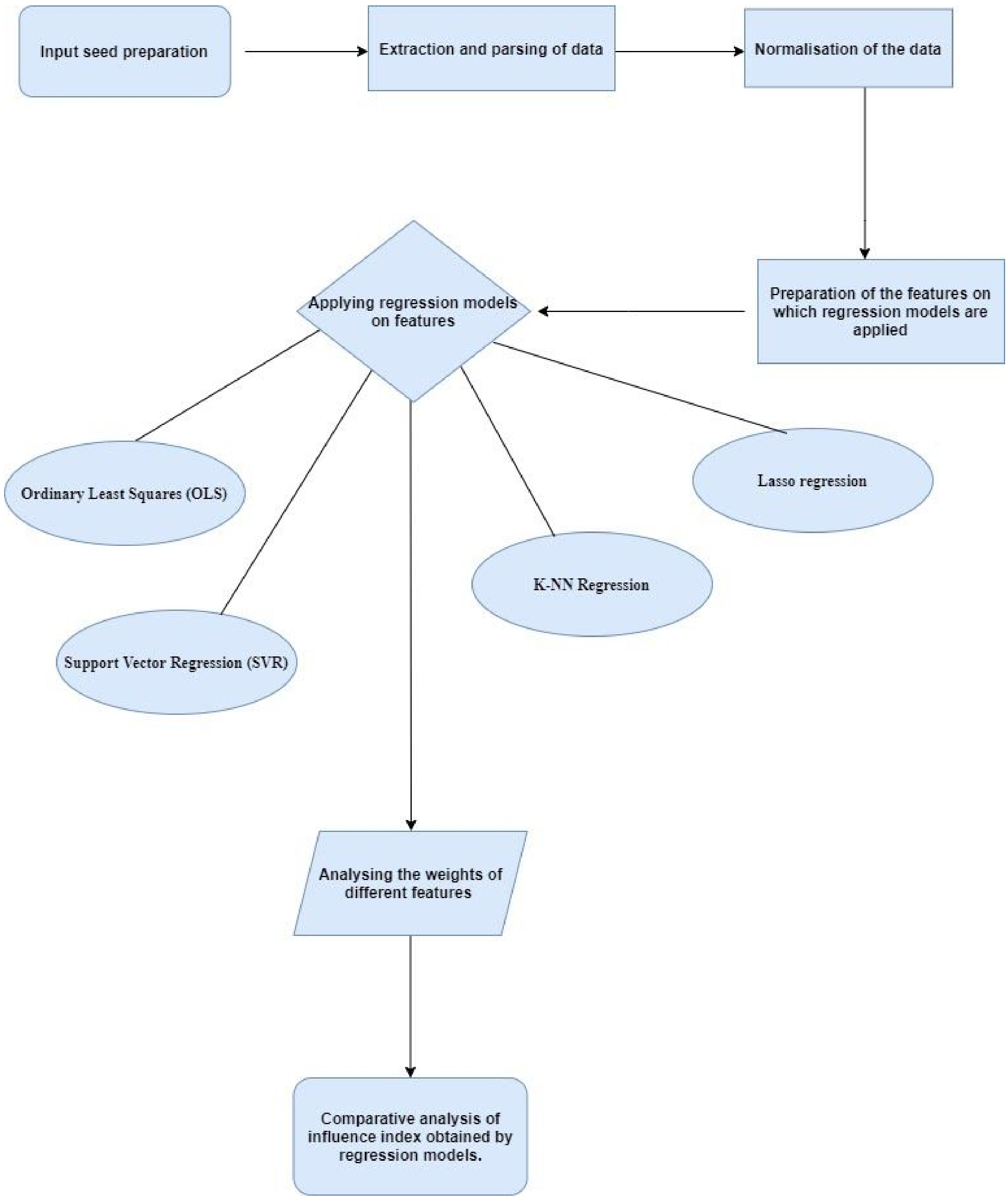

Fig. Workflow of project

BRIEF DESCRIPTION OF DATASETS

We have worked on Twitter dataset here with different features. The influencers' Twitter handles are used as the data extraction seed.These seeds are for a variety of celebrities from various fields, such as foreign athletes, media stars, and others.

First the developer account is made then data from Twitter is extracted via Tweepy and is collected and loaded in a CSV file. The CSV file is then processed in Pandas to extract and further process the twitter information. Several features from Twitter can be specifically used in modeling and research, whereas some other desired features are manually generated from the prevailing ones.

The table below shows the features to be used for our analysis.

Category	**Feature**	**Source**	**Definition**
Overall Footprint	Twitter Followers	Timeline	The total number of people who follow the brand's twitter account.
Engagement	Average Engagement	Tweet	This figure is calculated by averaging the amount of a tweet's retweets and favorites across all of the brand's tweets.
Outreach	Retweet	Post	The average number of retweets received by the brand's tweets
Hourly Engagement	Hour 1	Post	Total interaction received by posts on the site in the first hour after they were published.
	Hour 5	Post	Total interaction received by posts on the site in the 5th hour after they were published.

	Hour 10	Post	Total interaction received by posts on the site in the 10th hour after they were published.
Daily Engagement	Day 1	Post	Total engagement created by posts on the platform up to the first day after they were published.
	Day 2	Post	Total engagement created by posts on the platform up to the first day after they were published.
	Day 7	Post	Total engagement created by posts on the platform up to the first day after they were published.
Posting rate	Average post rate	Timeline	The frequency at which the individual tweets. This number is calculated by averaging the time interval between tweets.
Audience sentiment	Average Audience sentiment	Replies	For all of the tweets, the average meaning of overall sentiment was calculated using the bag of words method.

ALGORITHMS, TOOLS, AND TECHNOLOGIES TO BE USED

In our project we have analysed various features of the dataset. The regression algorithms used here are used to find the weights of the various features of the dataset. These are used to find the weights of the features so as to find out which feature affects more than others.

Linear Regression:

By fitting a linear equation to observed data, linear regression attempts to model the relationship between two variables. An explanatory variable is considered to be one variable, and a dependent variable is considered to be the other.

For regression problems, this capability, i.e., the expected performance of the trained function on unseen data, is typically assessed by measuring the MSE on a separate test set. Popular linear regression methods differ mainly in the choice of the regularization term. In particular, we consider four models:

1. Ordinary least squares (OLS) - In statistics, ordinary least squares (OLS) is a type of method for estimating unknown parameters in a linear regression model using linear least squares. OLS chooses the parameters of the linear function from a set of explanatory variables: the minimization of the sum of the squares of the differences between the observed dependent variable (the values of the observed variable) in the given dataset and those expected by the linear function. Mathematically $\Omega(w)=0$ (i.e., no regularization is used).
2. Support Vector Regression (SVR)- Support Vector Regression (SVR) is a regression technique that uses the same principles as SVM. It allows to model non-linear relationships between variables and therefore to tune hyperparameters to change the model's robustness. The aim is to essentially consider the points that are inside the decision boundary line. The hyperplane with the most points is the best fit line used for predictions. Contrary to OLS, the aim of SVR is to minimise the l2-norm of the coefficient vector.
3. K-NN Regression- KNN regression is a non-parametric approach that predicts the continuous variable by combining observations in the same neighborhood . For continuous variables Euclidean , Manhattan and Minkowski distances are used. For categorical variables, Hamming distance can be used. Examining the data first is the easiest way to determine the best value for K. A large K value is more accurate in general since it reduces overall noise.Although, it comes at the cost of blurring the distinct boundaries within the feature space.

4. LASSO - A type of linear regression that uses shrinkage is Lasso regression. Shrinkage is where data values, such as the mean, are reduced into a central point. Easy, sparse models (i.e. models with fewer parameters) are encouraged by the lasso method. The acronym "LASSO" stands for Operator of Least Absolute Shrinkage and Selection. Lasso regression performs L1 regularization, which adds a penalty equal to the absolute value of the magnitude of coefficients. Mathematically l1-norm regularization is used: $\Omega(w) = ||w1||$.

The models are compared on the basis of their mean Square Error, Mean Absolute Error, Root Mean Squared Error.

CONCLUSION

We have systematically analyzed the importance of Influence maximisation on social networks for digital marketing by various brands. The method to calculate the Social Influencer Index is studied and to be applied on the Twitter platform dataset. Data from twitter is extracted and different features are obtained after feature engineering and normalisation. After regression analysis ,the importance of different parameters will be measured to determine which contributes the most for influence maximisation. Finally the influencers are ranked on the basis of their influencing power to maximise the result of marketing strategies.

REFERENCES

1. *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.*
2. *Goodman, Michael B., Norman Booth, and Julie Ann Matic. "Mapping and leveraging influencers in social media to shape corporate brand perceptions." Corporate Communications: An International Journal (2011).*
3. *Keikha, Mohammad Mehdi & Rahgozar, Maseud & Asadpour, Masoud & Faghih Abdollahi, Mohammad. (2019). Influence Maximization across Heterogeneous Interconnected Networks based on Deep Learning. Expert Systems with Applications. 140. 112905. 10.1016/j.eswa.2019.112905.*
4. *Dwivedi, Yogesh K., et al. "Setting the future of digital and social media marketing research: Perspectives and research propositions." International Journal of Information Management (2020): 102168.*
5. *Chen, Wei, Yajun Wang, and Siyu Yang. "Efficient influence maximization in social networks." Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009*
6. *Liu, Shixi, et al. "Identifying effective influencers based on trust for electronic word-of-mouth marketing: A domain-aware approach." Information sciences 306 (2015): 34-52.*
7. *Singh, Shiwangi, Akshay Chauhan, and Sanjay Dhir. "Analyzing the startup ecosystem of India: a Twitter analytics perspective." Journal of Advances in Management Research (2019).*
8. *Cuevas, Leslie M., Sze Man Chong, and Heejin Lim. "Influencer marketing: Social media influencers as human brands attaching to followers and yielding positive marketing results by fulfilling needs." Journal of Retailing and Consumer Services 55 (2020): 102133.*
9. *Nafees, Lubna, and Christy M. Cook. "The impact of social media power on consumer attitudes toward the brand: The mediating/moderating role of social media source credibility." (2020).*
10. *Jie Tang, Jimeng Sun, Chi Wang and Zi Yang. Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009*
11. *K. Tago and Q. Jin "Tsinghua Science and Technology, vol. 23, no. 1, pp. 104-113, Feb. 2018"*
12. *Min Gao Li Xu Limei Lin Yanze Huang Xinxin Zhang Concurrency Computat Pract Exper. 2020;e5677*
13. *Lim, Xin Jean, et al. "The impact of social media influencers on purchase intention and the mediation effect of customer attitude." Asian Journal of Business Research 7.2 (2017): 19-36.*
14. *Author- M. S. Omar, A. Njeru and S. Yi. International Conference on Software Engineering, Artificial Intelligence,Networking and Parallel/Distributed Computing (SNPD), 2017, pp. 133-136*
15. *X. Zhou, B. Wu and Q. Jin. IEEE Transactions on Human-Machine Systems, vol. 48, no. 6, pp. 559-571, Dec. 2018*