Understanding And Clustering Hashtags According To Their Distributions Using Nlp In Machine Learning

S. SARAN KUMAR[1], S. SUHAIL BADHUUSHA [2], M. VISHWANATHAN[3], K. K. ARCHANA[4]

[1] *S. Saran Kumar, CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India.*
[2] *S. Suhail Badhuusha, CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India.*
[3]*M. Vishwanathan , CSE, Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India*
[4]*Mrs. K. K. Archana, Assistant Professor, CSE, Sri Ramakrishna Institute of Technology, Coimbatore, India*

Abstract:
The exponential increase of social media has caused a full-size quantity of user generated content often prepared by using hashtags powerful evaluation of hashtags can display trends person sentiment and different treasured insights however figuring out related hashtags and grouping them meaningfully gives a unique task this task introduces a hashtag clustering system using herbal language processing nlp strategies with long quick-term memory lstm neural networks at its core to analyze and cluster similar hashtags based on semantic similarity lstm a recurrent neural network version is nicely ideal for this project because of its capacity to seize contextual statistics inside sequences of text making it perfect for knowledge and classifying hashtags in social media contexts the gadget approaches a fixed of hashtags makes use of lstm to extract contextual and semantic relationships and applies clustering algorithms to group hashtags with comparable meanings or contexts to make the answer handy and consumer-pleasant an internet interface is evolved using flask a lightweight python net framework this interface allows customers to input a set of hashtags view clustered outcomes and explore the underlying semantic institutions the aggregate of lstm and flask enables an green interactive and scalable device for hashtag clustering with packages in fashion analysis advertising strategies and social media tracking

[1] INTRODUCTION

The explosive growth of social media structures such as twitter has given rise to the technology of a huge amount of unstructured information every day with more than 500 million tweets published every day twitter serves as an effective medium to capture the evolution of public opinion in real time and hashtags of ongoing discussions which can be specific key phrases or expressions following the symbol play a key function in organizing content on twitter allowing users to categorize posts according to specific topics and characteristics which reduces the difficulty of searching for related information but manual categorization is one of these large volumes of tweets using hashtags is challenging and the time spent on this challenge solves the task of predicting hashtags for tweets using devices to know predictive ways that mechanically assign hashtags can greatly improve statistical corporations and accessibility in social media systems these models are especially useful for marketing researchers and social scientists who depend on record categorization to analyze and influence sentiment trends hashtag prediction version uses nlp herbal language processing and device study strategies to research tweet content and predict by shooting the context of each tweet the final results is a simplified green machine that automatically matches usable hashtags to tweets improving statistics acquisition and trend analysis

1.1.Objectives

Content material management helps systems to create hashtags that decorate the discoverability of content material should you kindly rephrase the given text in a easy way moreover the version acts as a treasured tool for engaging in trend evaluation ensuring precision the usage of hashtag prediction can help analysts in monitoring the improvement of subjects and the sentiment of the general public the goal is to show off how system studying and natural language processing can automate the generation of hashtags in a easy way by using adapting a easy manner of writing that is contextually becoming and efficient we can beautify enterprise and coherence social media records that may be without difficulty searched.

1.2.Scope of the project

This projects scope consists of numerous critical ranges of gadget gaining knowledge of version improvement it begins with records series focusing on gathering a representative dataset of tweets that consists of each tweet text and related hashtags the subsequent degree involves facts preprocessing which prepares raw textual content facts by way of doing away with inappropriate characters eg urls punctuation tokenizing words and standardizing codecs to enhance consistency following preprocessing feature extraction techniques including tf-idf and word embeddings word2vec glove are applied to convert textual content into a numerical layout that the model can apprehend after preprocessing the version is educated using gadget getting to know algorithms ranging from easier fashions like logistic regression and svm to extra advanced methods like neural networks the model is evaluated on diverse

performance metrics accuracy precision bear in mind to make certain that it could expect hashtags appropriately the very last degree includes deployment planning with guidelines for integrating the version into social media structures or analytics equipment with the aid of completing these steps this venture pursuits to illustrate the feasibility of hashtag prediction and its capacity applications in social media management

1.3.Existing System

In modern social media and analytics platforms hashtag clustering and evaluation are regularly executed the use of basic text matching or key-wordbased totally methods those systems rely upon simple algorithms to perceive regularly happening hashtags grouping them based totally on genuine phrase matches or predefined categories whilst powerful for high-degree trend evaluation this technique lacks the capability to fully apprehend the context or semantic relationships between hashtags which limits its accuracy in clustering related tags a few superior systems use phrase embeddings or different herbal language processing nlp strategies to enhance hashtag analysis but many of those structures consciousness on static data and do now not adapt nicely to dynamic social media environments in which hashtags and traits shift swiftly these techniques additionally tend to rely upon statistical similarity alone missing out on the nuanced meaning that recurrent neural networks consisting of lstm can seize additionally few existing structures offer person-pleasant interfaces for actualtime hashtag clustering social media analysts or researchers often want to rely upon proprietary software or manually method facts in spreadsheets which may be time-consuming and inefficient while there are apis for trend evaluation theyll no longer provide clustering functions or require enormous technical know-how to put into effect this project addresses these gaps via using lstm fashions for semantic expertise of hashtags paired with a flaskbased internet utility for accessible real-time clustering and visualization

1.4. Proposed System

The proposed system introduces an intelligent hashtag clustering platform that uses lstm networks with longterm short-term memory for semantic analysis and contextbased hashtag clustering rather than just word matching by processing hashtags with the lstm model the system captures different meanings and relationships to enable accurate clustering of related hashtags although different in phrasing or structure the model is integrated into a flask-based web application that provides a user-friendly interface to work with real-time hashtag clustering and visualization users can input a list of hashtags and the system dynamically groups them into clusters based on contextual similarities obtained by the lstm model moreover without extensive technical knowledge this system is designed to adapt to evolving hashtag trends making it suitable for real-time applications in fastchanging social media environments this approach greatly improves the relevance and accuracy of hashtag clusters and supports more detailed analysis and trend identification.

1.5.Application

Hashtags can be clustered to become privy to trending subjects study public sentiment and understand the dynamics of online discourse companies leverage those insights for logo tracking aggressive assessment and optimizing marketing and advertising campaigns hashtag clustering additionally aids in identifying faux news or incorrect records tendencies. Nlp-based totally completely surely hashtag evaluation lets in producers intention particular audiences with the beneficial resource of the usage of manner of identifying generally co-taking region hashtags and their establishments with products sports or sentiments this allows for designing extra targeted and powerful promotional techniques. Hashtag clustering permits music traits in aim market possibilities show fan engagement and tailor content material material material fabric to particular person segments systems like netflix or youtube also can rent similar techniques to signify relevant content cloth cloth fabric cloth to customers.hashtag distributions can assist track conversations round precise sicknesses intellectual fitness issues or health tendencies this may additionally be prolonged to studying how wrong information or consciousness approximately illnesses eg covid-19 spreads at some stage in corporations

[2] LITERATURE SURVEY

Wei Feng et al., (2015) The research on streamcube for hierarchical spatiotemporal hashtag clustering provides valuable insights into the demanding situations and innovations of processing real-time social media statistics streams one major studying is the effectiveness of organizing hashtags into hierarchical records cubes based totally on time and space dimensions this technique enables customers to interactively explore twitter information with various granularities which include zooming into particular regions or intervals to discover localized or bursty activities it emphasizes the importance of dynamic scalable structures in the era of large records where real-time insights are crucial the improvement of a single-skip hashtag clustering set of rules showcases an efficient method for handling content materialevolving hashtags not like conventional clustering fashions streamcubes method permits for incremental updates ensuring that occasion clusters stay applicable as new information streams in this reinforces the necessity of adaptive algorithms that could accommodate the dynamic nature of social media content material an critical takeaway is the frameworks incorporation of event ranking mechanisms based on factors like reputation burstiness and localness those metrics provide a balanced way to become aware of and prioritize enormous occasions within precise

contexts whether worldwide or local this ranking gadget additionally highlights the flexibility of combining multiple rating elements to suit numerous packages from tracking tendencies to identifying emergent crises another key perception is the revolutionary use of a divide-andtriumph over approach to assemble and replace the streamcube which ensures excessive scalability and efficiency in processing big datasets the research underscores the potential of hybrid structures that merge spatial and temporal records aggregation to offer greater meaningful clustering and ranking consequences furthermore the take a look at highlights practical packages together with monitoring breaking information detecting natural screw ups and studying person sentiment at some stage in major events through leveraging geo-tagged tweets and temporal hierarchies the framework opens pathways for superior analytics in various fields from advertising and marketing to disaster response sooner or later the research identifies possibilities for future exploration including incorporating subject matter-based totally exploration and proactive alert mechanisms to enhance person engagement this work not best demonstrates the capability of streamcube but also sets the stage for broader improvements in social media analytics emphasizing the integration of actualtime records with user-interactive tools

Hsin-Min Lu, et al., (2015) I learned about the dynamics and intricacies of hashtag recommendation on Twitter, specifically focusing on the Topics-overTime Mixed Membership Model (TOT-MMM). This model builds on the premise that hashtags reflect evolving trends and events, often exhibiting distinct temporal patterns. TOTMMM uniquely captures these temporal clustering effects, enabling better modeling of time-sensitive latent topics in tweets. By treating text and hashtags as distinct yet complementary data types, it refines the prediction of hashtags by incorporating temporal and contextual elements. This approach improves upon prior methods such as similaritybased and LDA-based approaches, which faced challenges like data sparsity and inadequate consideration of temporal trends.The document elaborates on the application of TOT-MMM in identifying event-specific hashtags and its evaluation against baseline models using extensive datasets. Its combination with token-level similarity approaches enhanced performance, underscoring the value of integrating latent-topic-based features with surfacelevel features. Beyond recommendation, the model demonstrated utility in event signal detection, as evidenced by simulations around the BP oil disaster. TOTMMM effectively detected hashtags tied to this event with improved timeliness and accuracy compared to other models. The findings illuminated the importance of temporal features and hybrid modeling in refining social media analytics.

Giovanni Stilo, et al., (2016) addressed the challenges in analyzing hashtags on social media mainly twitter due to issues like polysemy multiple meanings and synonymy unique hashtags representing the equal concept these problems make it hard to institution tweets by means of subject matter seeing that hashtags range widely across languages and arent without problems classified to address this the authors introduce sax a clustering set of rules that companies hashtags based on their temporal usage styles via examining how hashtags co-occur and vary over the years sax businesses related hashtags that show off comparable temporal behaviors like spikes in utilization at some stage in precise occasions sax makes use of symbolic combination approximation to convert hashtag frequency into symbolic representations permitting clustering of similar usage patterns with out relying on lexical content material this temporal approach outperforms conventional strategies that depend on lexical similarity which struggles with twitters short and sundry content material through widespread checking out sax is shown to successfully capture related hashtags for occasions demonstrating higher semantic cohesion and computational efficiency than existing strategies the paper concludes that temporal similarity is a strong criterion for clustering hashtags providing a capability utility in regions like fashion analysis and hashtag recommendation
structures

Bénédicte Le Grand, et al., (2017) i found out approximately a unique approach for hashtags and user pointers on twitter leveraging semantic evaluation and clustering techniques hashtags as metadata labels play a crucial role in indexing and categorizing tweets thereby enhancing consumer revel in and facts accessibility this study proposed a strong recommender system architecture that includes semantic evaluation to enhance social user profiles and employs clustering to organization users with shared pastimes key takeaways include the introduction of two modern ranking schemes hf-iutf and hf-ighf the hfiutf scheme refines consumer profiles by means of ranking hashtags based totally on their frequency in a users tweets and their distinctiveness across all tweets the hf-ighf scheme identifies consultant hashtags for clusters of similar users primarily based on their usage frequency in the organization and their relative popularity these rating techniques stimulated via tf-idf concepts beautify the precision of pointers by means of accounting for temporal and contextual co-occurrences of hashtags the structures pipeline starts with reading person tweets indexing hashtags and filtering beside the point content material semantic enrichment is accomplished by way of grouping hashtags with similar meanings and augmenting them with associated terms the usage of assets like wordnet this twinlayered semantic technique guarantees profiles capture a complete and nuanced expertise of consumer interests furthermore temporal and contextual analyses pick out hashtags that regularly co-arise in tweets or inside carefully timed postings including intensity to the recommendation procedure spectral clustering became carried out to institution users with comparable profiles allowing tips for users to observe or hashtags to undertake primarily based on shared pursuits this clustering-based advice is flexible

extending to any similarity grouping mechanism the architectures novelty lies in integrating semantic analysis with clustering and its independence from specific clustering methodologies overall this method advances the kingdom of customized guidelines on twitter shifting beyond conventional strategies by way of blending semantic insights with sophisticated ranking and clustering techniques the have a look at highlights the significance of wealthy user profiles semantic relationships and dynamic ranking strategies in enhancing social media hints even as additionally setting the degree for in addition research on implementing and benchmarking such structures
towards present solutions

Nartlada Bhakdisuparit, et al., (2018) targeted on clustering hashtags by using reading word distributions to capture trending subjects in actual time the authors acquire sample tweets and extract hashtags then gather tweets associated with each hashtag and generate word distributions for analysis to measure similarity they use jensen-shannon divergence between phrase distributions of hashtags which allows clustering based totally on content similarity the method includes several steps gathering and filtering tweets extracting hashtags computing word frequency distributions and clustering based on jensen-shannon divergence 8 this outcomes in a hierarchical dendrogram showing associated hashtag organizations the studys findings consist of predominant clustersone on american politics and any other on track awards this clustering framework pursuits to gain advertising techniques by means of better focused on consumer pastimes primarily based on trending topics future paintings will cope with unsolicited mail filtering and consumer-based totally hashtag categorization

Vladimir Barash, et at., (2018) i won an information of ways spectral clustering may be carried out to analyze hashtag adoption behaviors on twitter focusing on interestbased totally network detection the studies centered across the bringbackourgirls marketing campaign exploring how clustering methods ought to organization customers with shared interests and predict destiny hashtag adoptions spectral clustering making use of consumer adoption information validated efficiency in revealing communities with aligned topical engagements not like clustering primarily based on social network ties it furnished greater granular insights into person pastimes surpassing network-based clusters in predicting topical behaviors i learned about the benefits of spectral clustering over conventional strategies like latent dirichlet allocation lda in phrases of computational efficiency although lda provided barely higher predictive accuracy moreover the contrast with the louvain technique highlighted spectral clusterings superiority in generating clusters that align better with actual hobby companies specially whilst carried out to artificial datasets the integration of unsupervised techniques to discover person groups and assess their behavior regarding new hashtags become particularly insightful showcasing how such techniques can offer predictive insights into social moves dynamics and person engagement styles the analysis emphasized the interaction of algorithm performance accuracy and the importance of leveraging adoption styles to discern and are expecting hobby-based totally network behaviors efficaciously

Nada Ben-Lhachemi, et al., (2018) evolved a tool for recommending relevant hashtags to twitter clients the authors propose the usage of word embeddings and clustering to capture 9 semantics within tweets and provide correct hashtag recommendations on this tweets are represented the use of a weighted commonplace of word embeddings from a preprofessional model google information word2vec this technique captures each syntactic and semantic context inside tweets.

El Habib Nfaoui et al., (2018) the studies on streamcube for hierarchical spatiotemporal hashtag clustering presents precious insights into the demanding situations and improvements of processing actual-time social media records streams one important studying is the effectiveness of organizing hashtags into hierarchical statistics cubes based on time and area dimensions this approach enables users to interactively discover twitter data with varying granularities which includes zooming into unique regions or durations to discover localized or bursty activities it emphasizes the importance of dynamic scalable systems within the generation of huge facts where actual-time insights are essential the improvement of a unmarried-skip hashtag clustering algorithm showcases an green technique for coping with content material-evolving hashtags unlike traditional clustering models streamcubes method allows for incremental updates ensuring that event clusters continue to be relevant as new information streams on this reinforces the necessity of adaptive algorithms which could accommodate the dynamic nature of social media content an crucial takeaway is the frameworks incorporation of event ranking mechanisms based totally on factors like popularity burstiness and localness those metrics offer a balanced manner to perceive and prioritize good sized events inside particular contexts whether or not global or nearby this ranking system also highlights the flexibility of combining multiple ranking factors to in shape numerous packages from monitoring developments to figuring out emergent crises any other key perception is the revolutionary use of a divide-and-overcome approach to construct and update the streamcube which ensures high scalability and performance in processing big datasets the research underscores the potential of hybrid systems that merge spatial and temporal facts aggregation to provide extra meaningful clustering and ranking effects moreover the take a look at highlights realistic packages including tracking breaking information detecting natural screw ups and reading user sentiment throughout foremost activities by means of leveraging geo-tagged tweets and temporal hierarchies the framework opens pathways for superior analytics in numerous fields from advertising to catastrophe response eventually the studies identifies opportunities for future exploration inclusive of

incorporating subject matter-based totally exploration and proactive alert mechanisms to beautify consumer engagement this paintings no longer handiest demonstrates the capability of move dice however also sets the stage for broader improvements in social media analytics emphasizing the mixing of real-time records with consumer-interactive gear

Mahmoud Rokaya, et al., (2021) improved method for clustering hashtags on social media by way of combining textual content-based totally lexical clustering semantic metadata clustering and a new enhancement known as power hyperlinks hashtag clustering is critical for content material categorization permitting social media platforms to organization related topics more efficaciously however usergenerated content material often consists of spelling and grammar inconsistencies that could lessen the effectiveness of widespread clustering techniques historically hashtag clustering uses either lexical strategies which analyze the text homes of hashtags or semantic techniques which use external metadata resources like wordnet and wikipedia to define relationships between hashtags but both methods have limitationslexical clustering can misclassify hashtags when consumer language is inconsistent and semantic strategies may struggle without metadata to cope with those gaps this research introduces a hybrid method that merges lexical and semantic methods however adjusts weighting based totally on hashtag distance to each cluster center resulting in a more delicate clustering process additionally the observe integrates strength links which leverage co-prevalence patterns to similarly refine clusters through considering how regularly words and hashtags seem together strength hyperlinks serve to reallocate hashtags greater correctly by using recognizing institutions primarily based on their frequency of appearing in similar contexts this hybrid method greater through electricity links has been fastidiously examined using both a managed dataset and a large random pattern from twitter consequences display that the new technique outperforms conventional clustering strategies enhancing the consistency and relevance of hashtag groupings the findings have sensible implications for social media analytics junk mail discount sentiment evaluation and search engine optimization by enhancing the clustering accuracy the approach 7 supports greater nuanced content material type permitting better consumer studies and greater precise subject matter corporation.

3. REQUIREMENT SPECIFICATIONS

SOFTWARE DESCRIPTION WAMP Server

WAMPs are collections of separately developed apps that are installed on Windows-based PCs. A computer network, such the internet or a private network, can serve dynamic web pages thanks to the interaction of these applications. LAMP is the name of the corresponding installation on a Linux operating system. MAMP is the comparable installation for the Mac operating system. SAMP is the name of the equivalent installation on a Solaris operating system. FAMP is the name of the analogous installation on a FreeBSD operating system. The initials of the operating system (Windows) and the package's main constituents—Apache, MySQL, and PHP (or Perl or Python)—are combined to make the acronym "WAMP".

APACHE

One web server that stands out for having been crucial to the early development of the World Wide Web is the Apache HTTP Server, also known as just Apache. Linux web servers make up the vast bulk of all web servers that use Apache.

Under the direction of the Apache Software Foundation, an open development community creates and maintains Apache. Numerous operating systems, including as UNIX, FreeBSD, Linux, Solaris, Novell NetWare, Mac OS X, Microsoft Windows, OS/2, TPF, and eComStation, are compatible with the application. Apache is classified as free and open source software, and it was released under the Apache License.

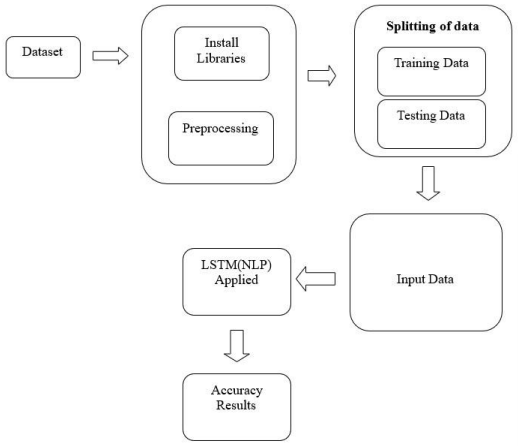
MySQL

With over 11 million installations, MySQL is a relational database management system (RDBMS). The application functions as a server that grants several users access to various databases.The Swedish corporation MySQL AB, which is currently a division of Sun Microsystems, is the owns the copyright to the majority of the software. The GNU General Public sole owner and sponsor of MySQL. It also License and various proprietary agreements govern the availability of the project's source code.

4.METHODOLOGY

LSTM is a kind of rnn used for managing continueos statistics and seize prolonged-term dependency in evaluation to preferred rnns lstms address the vanishing gradient problem via incorporating a memory mobile and 3 gates input neglect approximately and output the ones gates regulate the float of statistics finding out what to characteristic eliminate or bypass to the following americathe memory cellular continues applicable information through the years allowing lstms to version complicated longvariety styles widely utilized in responsibilities like natural language processing speech recognition

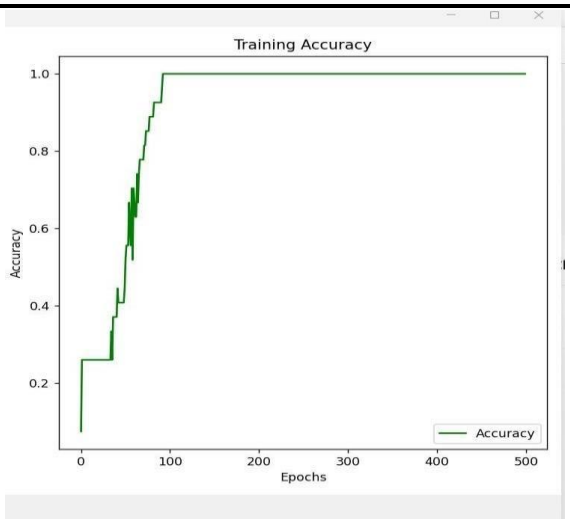
4.2. BLOCK DIAGRAM



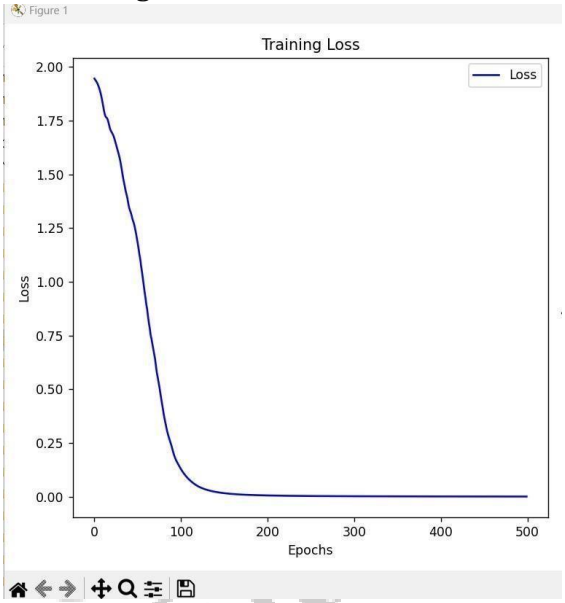
5.RESULTS

5.1 Training accuracy

This shows the schooling accuracy progression of a system learning version over 500 epochs the x-axis tells the range of epochs even as the y-axis tells the accuracy finished all through education at the beginning the accuracy is low indicating that the fashions initial predictions arent very reliable as training progresses the accuracy increases considerably mainly at some stage in the primary one hundred epochs reflecting that is mastering from the information and enhancing its predictive capabilities the pointy upward thrust in accuracy throughout the early ranges highlights the effective optimization of the models parameters after about one hundred epochs the accuracy methods 10 or one hundred signifying that the version has nearly perfectly learned the training statistics this flat line toward the give up shows that the model has done maximum accuracy and has stabilized such a trend is acceptable in training however ought to be evaluated carefully whilst excessive schooling accuracy demonstrates the models capacity to in shape the schooling data it is crucial to validatetheversion



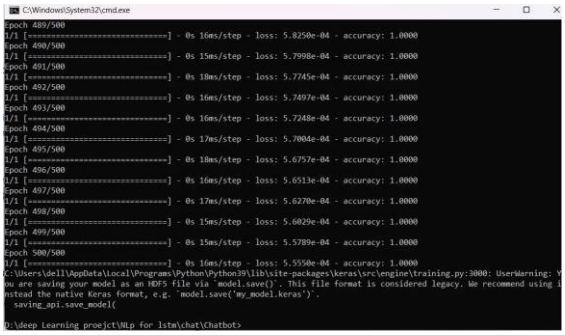
5.2 Training loss



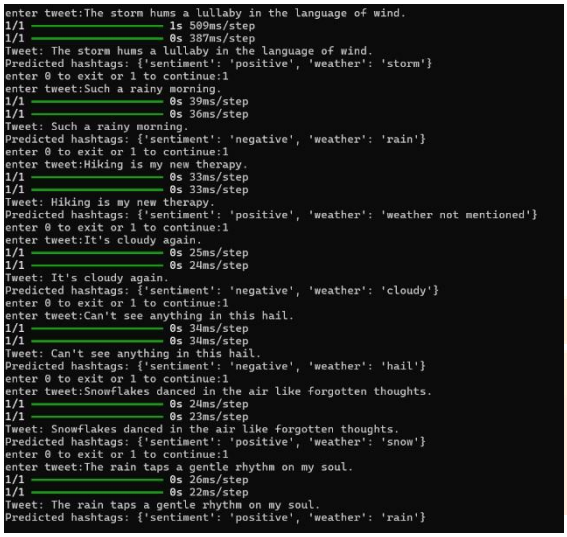
5.3 Training the model

The figure displays the terminal output from versions process the version are trained over 500 epochs as indicated by way of the epoch counter the key metrics located include loss and accuracy the loss value regularly decreases as the schooling progresses achieving approximately 5555e-04 inside the five hundredth epoch this shows that the models blunders in predicting outputs has minimized notably the measure which evaluates the versions overall scores at the schooling dataset achieves an ideal rating of one000 100 via the very last epoch this signifies that the version correctly discovered the patterns within the schooling data to make accurate predictions for all instances moreover the log consists of a caution message related to the versions saving method it recommends saving the skilled model inside the updated keras format keras rather than the older hdf5 layout h5 as the latter is taken into consideration legacy this output demonstrates a successful education manner reaching ideal outcomes with minimum loss and best accuracy but such high schooling accuracy necessitates similarly evaluation the

usage of test dataset to make sure it generalizes properly and avoids overfitting



5.4 Output



6. Conclusion

This task successfully demonstrates a strong technique to hashtag clustering the use of nlp techniques and lstm fashions offering an revolutionary tool for studying social media trends by leveraging the strengths of lstm networks in shooting contextual relationships the device can correctly institution hashtags with semantic similarity even when they vary in phrasing or structure the integration with a flaskbased net application guarantees a person-pleasant experience enabling customers to have interaction with the version effortlessly and advantage insights in real time this clustering answer addresses obstacles of conventional word-based clustering by means of focusing on context which notably enhances the relevance of clusters for entrepreneurs researchers and social media analysts the adaptability of the model to new developments in addition supports continuous relevance in dynamic social media landscapes destiny enhancements ought to consist of expanding the versions talents to identify rising subjects robotically and applying it to different nlp-based clustering programs broadening its application in social media analytics and past.

7.REFERENCES

[1] Orianna Demasi, Douglas Mason and Jeff Ma, "Understanding Communities via Hashtag Engagement: A clustering Based Approach," in Proc. ICWSM, 2016

[2] Tasuku Kimura and Miyamori Hisashi, "A Method of classifying Relationships between Hashtags Using Cooccurrence and Latent Topics," in IEICE Transactions on Information and Systems (Japanese Edition) Vol.J98-D No.8 pp.1151-1161, 2015.

[3] Hiroya Takamura, Introduction to Machine Learning for Natural Language Processing (in Japanese), Corona Publishing Co.,LTD, Tokyo, Jpn, 2015.

[4] Yuta Noro, Masaharu Hiroshi, Yokoyama Shouhei and Hiroshi Ishikawa, "Proposal of a System for Auomatically Hashtagging in Microblog," in DEIM Forum, 2013

[5] Sandjai Bhulai, Peter Kampstra, Lidewij Kooiman, Ger Koole, Marijin Deurloo and Bert Kok, "Trend Visualization on Twitter: What's Hot and What's Not?," in DATAANALYTICS, 2012.

[6] Cristina I. Muntean, Gabriela A. Morar and Darie Moldovan, "Exploring the Meaning behind Twitter Hashtags through Clustering," in International Workshops and future Internet Symposium BIS2012, 2012.

[7] Zangerle E, Gassler W, Specht G, On the Impact of Text Similarity Functions on Hashtag Recommendations. Microblogging Environments in Social Network Analysis and Mining Springer. DOI: 10.1007/s13278- 013-0108-x, 2013

[8] M. Michelson and S. A. Macskassy, "Discovering users' topics of interest on twitter: A first look," in Proceedings of the 4th Workshop on Analytics for Noisy Unstructured Text Data, ser. AND '10. New York, NY, USA: ACM, 2010, pp. 73–80. [Online]. Available: http://doi.acm.org/10.1145/1871840.1871852

[9] Y. Shi and M. Macy, "Measuring structural similarity in large online networks," Social Science Research, vol. 59, pp. 97–106, 2016.

[10] She J, Chen L, TOMOHA: TOpic model-based HAshtag recommendation on twitter, In: Proceedings of the 23rd International Conference on World Wide Web, 2014

[11] Gong Y, Zhang Q, Hashtag Recommendation Using Attention-Based Convolutional Neural Network, In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016

[12] KALLOUBI, Fahd et NFAOUI, El Habib. Learning to Suggest Hashtags: Leveraging Semantic Features for Time-Sensitive Hashtag Recommendation on theTwitter Network. In: Semantic Web Science and RealWorld Applications. IGI Global, 2019

[13] Y Gong, Q Zhang, Hashtag Recommendation Using Attention-Based Convolutional Neural Network, InProceedings of the Twenty-Fifth International Joint
Conference on Artificial Intelligence, IJCAI-16,2016