A Major Project Report On

#### HASHTAG PREDICTION FOR SOCIAL MEDIA

Submitted in fulfillment of the requirements for the award of the

#### Bachelor of Technology

In

#### Department of Computer Science and Engineering

##### By

**GV Harshith 20241A05E7**

**A Sahith 20241A05C4**

**B Shiva Krishna 20241A05D6**

**K Ravichandra 21245A0516**

**A Guna Teja 20241A05C1**

Under the Esteemed guidance of

#### D USHA SREE

#### (Assistant Professor, CSE Department)



**Department of Computer Science and Engineering**

**GOKARAJU RANGARAJU INSTITUTE OF ENGINEERING AND TECHNOLOGY**

**(Autonomous)**

**Bachupalli,Kukatpally,Hyderabad,Telangana,India,50090**

**2023-2024**



## GOKARAJU RANGARAJU

**INSTITUTE OF ENGINEERING AND TECHNOLOGY**

**(Autonomous)**

## **CERTIFICATE**

This is to certify that the major project entitled “**Hashtag Prediction for Social Media**” is submitted by **GV. Harshith(20241A05E7), A. Sahith(20241A05C4), B. Siva Krishna(20241A05D6), K. Ravichandra(21245A0516), A. Guna Teja(20241A05C1),** in fulfillment of the award of a degree in BACHELOR OF TECHNOLOGY in Computer Science and Engineering during the academic year **2023-2024.**

INTERNAL GUIDE HEAD OF THE DEPARTMENT

**D Usha Sree Dr. B. SANKARA BABU**

**Assistant Professor Professor**

**EXTERNAL EXAMINER**



Many people helped us directly and indirectly to complete our project successfully. We would like to take this opportunity to thank one and all. First, we wish to express our deep gratitude to our internal guide **D Usha Sree**, **Assistant Professor**, Department of CSE for his support in the completion of our project report. We wish to express our honest and sincere thanks to **Dr. B. Sankara Babu, HOD,** Department of CSE, and to our principal **Dr. J. Praveen** for providing the facilities to complete our major project. We would like to thank all our faculty and friends for their help and constructive criticism during the project completion phase. Finally, we are very much indebted to our parents for their moral support and encouragement to achieve goals.

**GV Harshith 20241A05E7**

**A Sahith 20241A05C4**

**B Shiva Krishna 20241A05D6**

**K Ravichandra 21245A0516**

**A Guna Teja 20241A05C1**

## **DECLARATION**

We hereby declare that the major project entitled “**Hashtag Prediction for Social Media**” is the work done during the period from **2023-2024** and is submitted in the fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering from **Gokaraju Rangaraju Institute of Engineering and Technology (Autonomous under Jawaharlal Nehru Technology University, Hyderabad).** The results embodied in this project have not been submitted to any other university or Institution for the award of any degree or diploma.

**GV Harshith 20241A05E7**

**A Sahith 20241A05C4**

**B Shiva Krishna 20241A05D6**

**K Ravichandra 21245A0516**

**A Guna Teja 20241A05C1**

|  |  |  |
| --- | --- | --- |
|  | Table of Contents |  |
| Chapter | TITLE | Page No |
|  | Abstract | 1 |
| 1 | Introduction | 2-3 |
| 2 | System Requirements | 4 |
|  | 2.1 Software Requirements | 4 |
|  | 2.2 Hardware Requirements | 4 |
|  | 2.3 Data Set | 4 |
| 3 | Literature Survey | 5 |
| 4 | Propose Approach , Modules Description, and UML Diagrams | 6-9 |
|  | 4.1 Modules | 6 |
|  | 4.2 UML Diagrams | 7-9 |
| 5 | Implementation, Experimental Results &Test Cases | 10-13 |
| 6 | Conclusion and Future Scope | 14 |
| 7 | References | 15 |
|  | Appendix   1. Full Paper-Publication Proof 2. Snapshot Of the Result 3. Optional (Like Software Installation /Dependencies/ Sudo code) | 16 |

|  |  |  |
| --- | --- | --- |
|  | **LIST OF FIGURES** |  |
| **Fig No** | **Fig Title** | **Page No** |
| Fig-4.2.1 | System Process | 7 |
| Fig-4.2.2 | Input and Preprocess | 8 |
| Fig-4.2.3 | Testing | 9 |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  | **LIST OF TABLES** |  |
| **Table No** | **Table Name** | **Page No** |
| 5 | Test Cases | 13 |

**ABSTRACT**

In the age of social media, the use of hashtags has become a prevalent way for users to categorize and label their posts, facilitating content discovery and engagement. However, the manual selection of relevant hashtags for a post can be time-consuming and error-prone. To address this challenge, hashtag prediction has emerged as a valuable task in the realm of natural language processing and machine learning. The goal of hashtag prediction is to automatically suggest a set of relevant hashtags for a given piece of text, enhancing post visibility and engagement. In this project, we will create a Machine Learning model. The model will take tweets as input and separate each word of those tweets into three groups, namely “Sentiments”, “When,” and “Kind.” In these types, first of “Sentiments” will describes how the people feel about the things and second “When” category is used to find out the past, present and future from the sentence. Third category of “Kind” is used to describe the type of situation has happened. For example, weather conditions like ‘Sunny’, ‘Rainy’ , ‘Stormy’ , ‘Lightning’ are described in this category.

# **1 INTRODUCTION**

The capacity to precisely forecast popular hashtags has become critical for influencers, marketers, and users alike in the dynamic world of social media. Hashtags are an effective tool for audience interaction, content discovery, and online visibility. The difficulty, though, is predicting which hashtags will catch on and appeal to users immediately. Predictive analytics can help in this situation by using machine learning algorithms and data-driven insights to estimate hashtag popularity and relevance ahead of time. Predictive algorithms can identify new trends and recommend the best hashtags for optimizing reach and engagement by examining past data, user behaviour patterns, and contextual factors. With the amount of content being posted on social media sites growing, hashtag prediction gives businesses a competitive edge. The model will take tweets as input and separate each word of those tweets into three groups, namely “Sentiments”, “When,” and “Kind.”

In the present framework, we demonstrate the superior performance in the method and compares with the existing unsupervised approaches like Word2Vec and supervised methods like WSABIE for hashtag prediction tasks. In our existing system, we are using Many to Many Encoder-Decoder Sequence model. This model accepts the stream of sentences as inputs and outputs another stream of sentences. The main techniques used in this model are encoder and decoder along with the neural network layers. In the encoder model, it takes the input of text was given by user and passes it to 3 stacked layers of Long Short-Term Memory (LSTM) network and after processing of the text then returns the hidden state and passes to the next sequence using one LSTM layer and Attention mechanism. This mechanism mainly helps to model to focus only on specific words for the hashtag prediction.

The main advantages of this existing system will indicate its efficiency in capturing and understanding semantic relationships within textual data. It shows the versality by transferring successfully to the personal document tasks that are recommended. It suggests the system capabilities extended beyond its original training tasks and this system makes it applicable to diverse applications. This trained on hashtags prediction consistently outperform those supervised learning methods and indicates the robustness and reliability of the system of various evaluation scenarios. This system architecture exhibits superior performance compared to others and even when both are trained on the same predictions task. This suggests that the chosen architecture shows significantly to the model effectiveness.

# **2 SYSTEM REQUIREMENTS**

# **2.1 Software Requirements:**

Operating System : Windows or Mac OS

Platform : VS code, IDLE

Programming Language : Python

Frontend Library : ReactJS

# **2.2 Hardware Requirements:**

1. CPU or GPU

2. Memory (RAM)

3. Storage

4. Parallelization

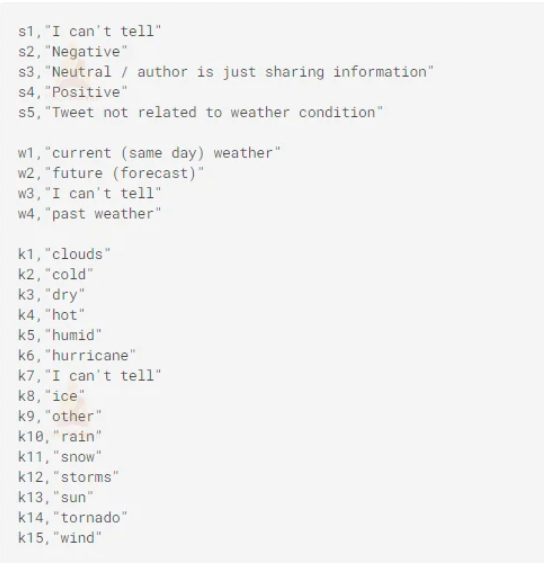
5. Optimization Techniques

6. Inference Hardware

7. Monitoring and Scaling

# **2.3 Data Set**

<https://www.kaggle.com/c/crowdflower-weather-twitter/data>



# **3 .LITERATURE SURVEY**

This Hashtag prediction for social media project, existing literature key insights into this evolving field. The goal of this existing system is to streamline and enhance the user experience by simplifying the hashtag selection process, and ultimately to boost up the post visibility and engagement. The Researcher have increasingly focused on leveraging natural language for textual content and several seminal studies reaches into specific aspects of prediction. In our project, the system will understand the emotional aids in suggesting hashtags that align with overall moods, sentiments expressed by the content. In this project, we used to generate the hashtags based on three categories which are “Sentiments”, “When”, “Kind”. So, these categories will investigate the other significant area. By making judgement on the underlying theme or topic of text, these models contribute to more accurate and generate relevant hashtags.

We aim to contribute this body of knowledge in the project by designing a ML model for hashtags prediction on tweets. Our model approach is to categorized each word in the tweets into three categories: ‘Sentiment’ for emotional hashtags, ‘When; for temporal relevance hashtags, and ‘Kind’ for content hashtags. This approach draws inspiration from and builds upon the insights from existing literature and our model aligns with best practices and latest advancements in hashtag prediction. Through this comprehensive literature survey, we need to understand the current state of research but also to explain meaningfully to evolve the field of hashtag prediction.

# **4 .PROPOSED APPROACH**

# **4.1 MODULES**

**4.1.1 Module-1: Input and Preprocessing**

In our system, users can provide the input text directly and with both input types undergoing processing techniques. When the system takes the text as input then it preprocesses the text to standardize, its format and improve analysis accuracy before that it analyzes across the text into three dimensions are sentiment, when and kind. Our system utilizes the machine learning algorithms trained on labeled datasets, this model generates hashtags based on the content given by user in sentiments, temporal context, topic. This system finally predicts hashtags users and optimize their social media presence effectively.

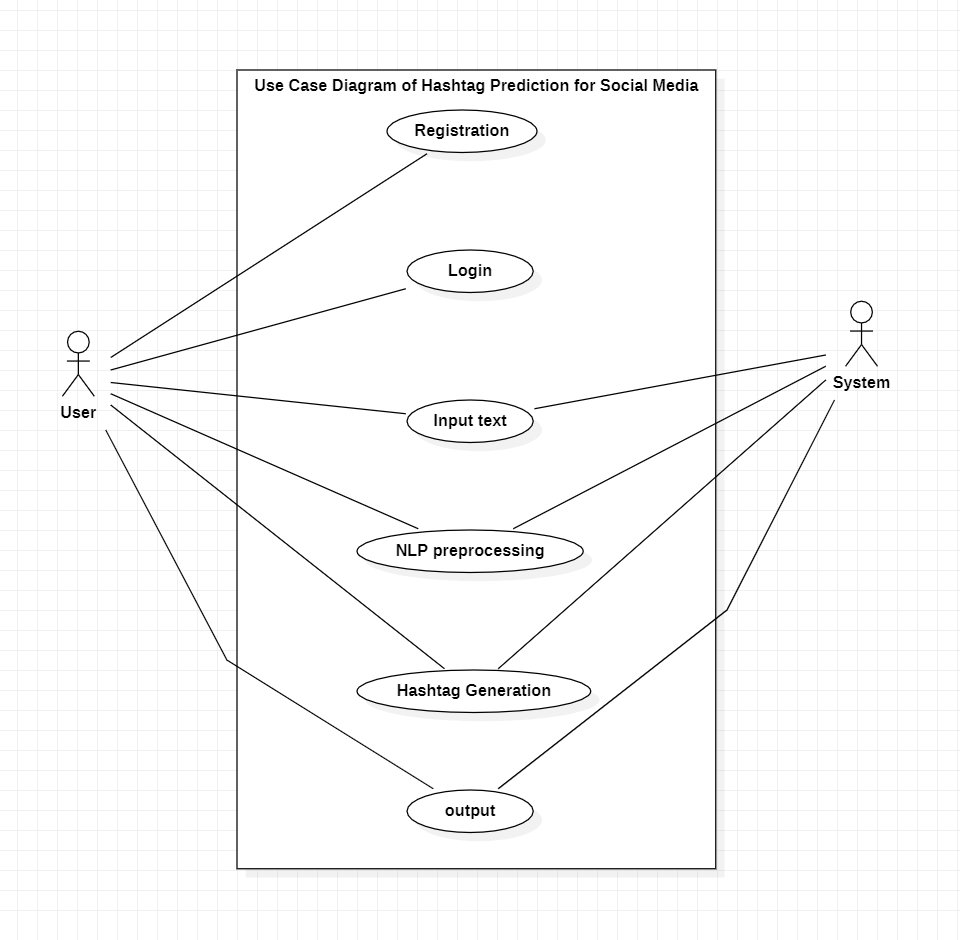
**4.1.2 Module-2: Testing**

This systemanalyzes the user input across various dimensions, including sentiments, temporal, and topic. Then the input text will be preprocessed then the system extracts relevant features fromthe text. These features capture the input content, enabling our algorithm to predict the relevant hashtags. Our system delivers accurate and tailored hashtags based on the content given by user. With the help of our system, user can optimize the presence on the social media and connect their audience with more effectively.

# **UML DIAGRAMS**

**1.UML Diagram**

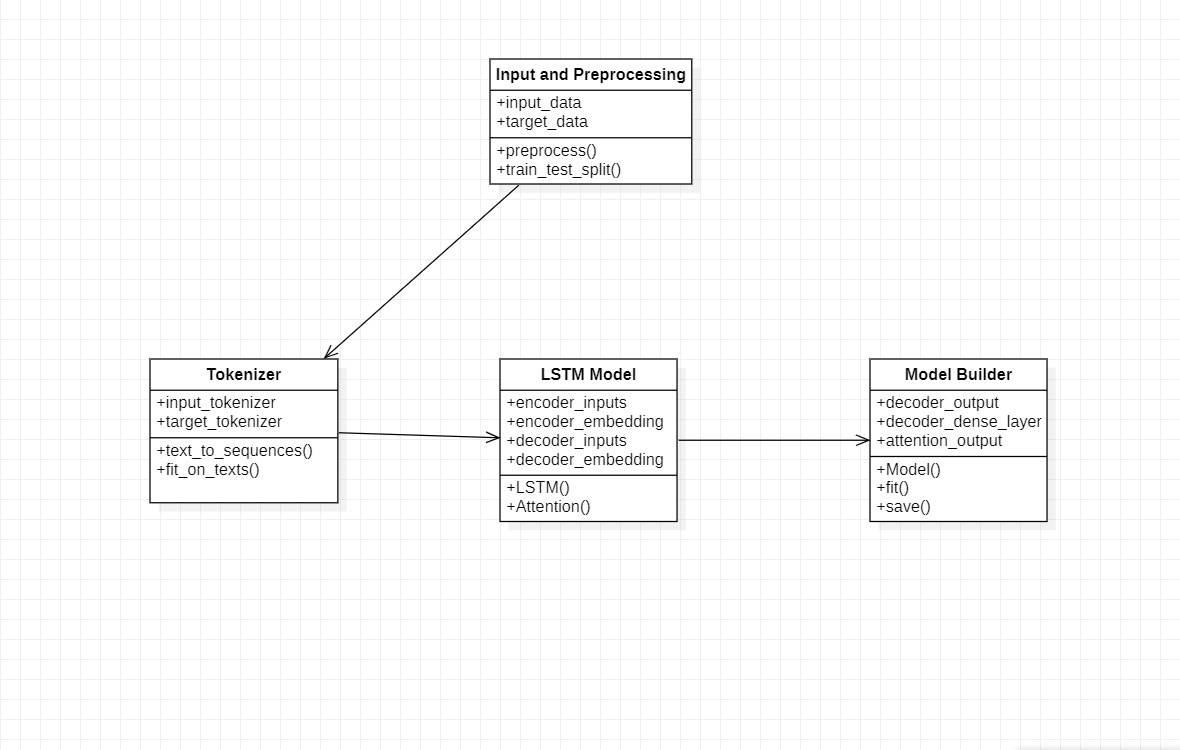
This diagram, is a way to visualize systems and software using UML software create UML diagrams to understand the designs, architecture, and proposed implementation of complex software systems. The diagrams are also used to model workflows and business. The Unified Modeling Language (UML) simplifies the information to visual reference that’s easier to digest. The aim of UML, is to define a standard way to visualize a way of system has been designed. It is quite similar to blueprints used in other fields of engineering.



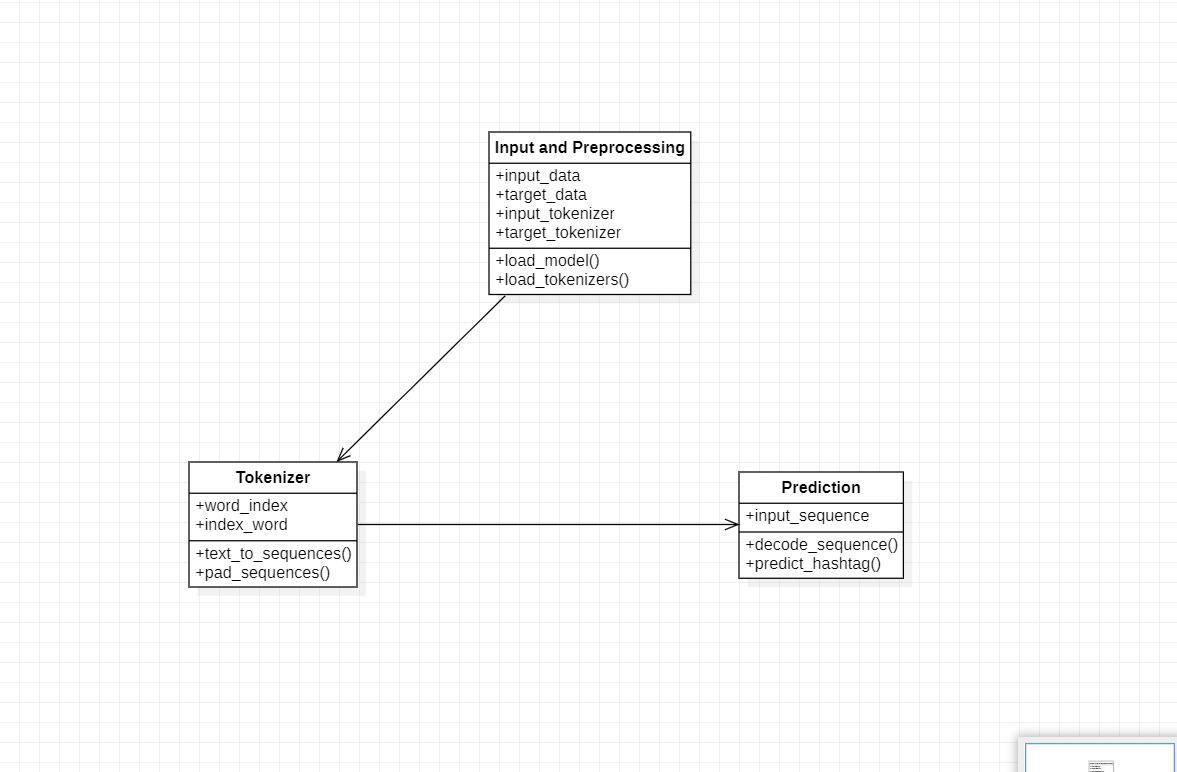
**Figure-4.2.1: System Process**

**2. Class Diagram**

The Class Diagram is a type of UML diagram used in software engineering to visually represent the structure and relationships of class inn system. It is used to represent the structure and relationships of class within system which is used to construct the object-oriented system. It also used to visualize the object-oriented system and it helps to communicate and document in the structure of the software.

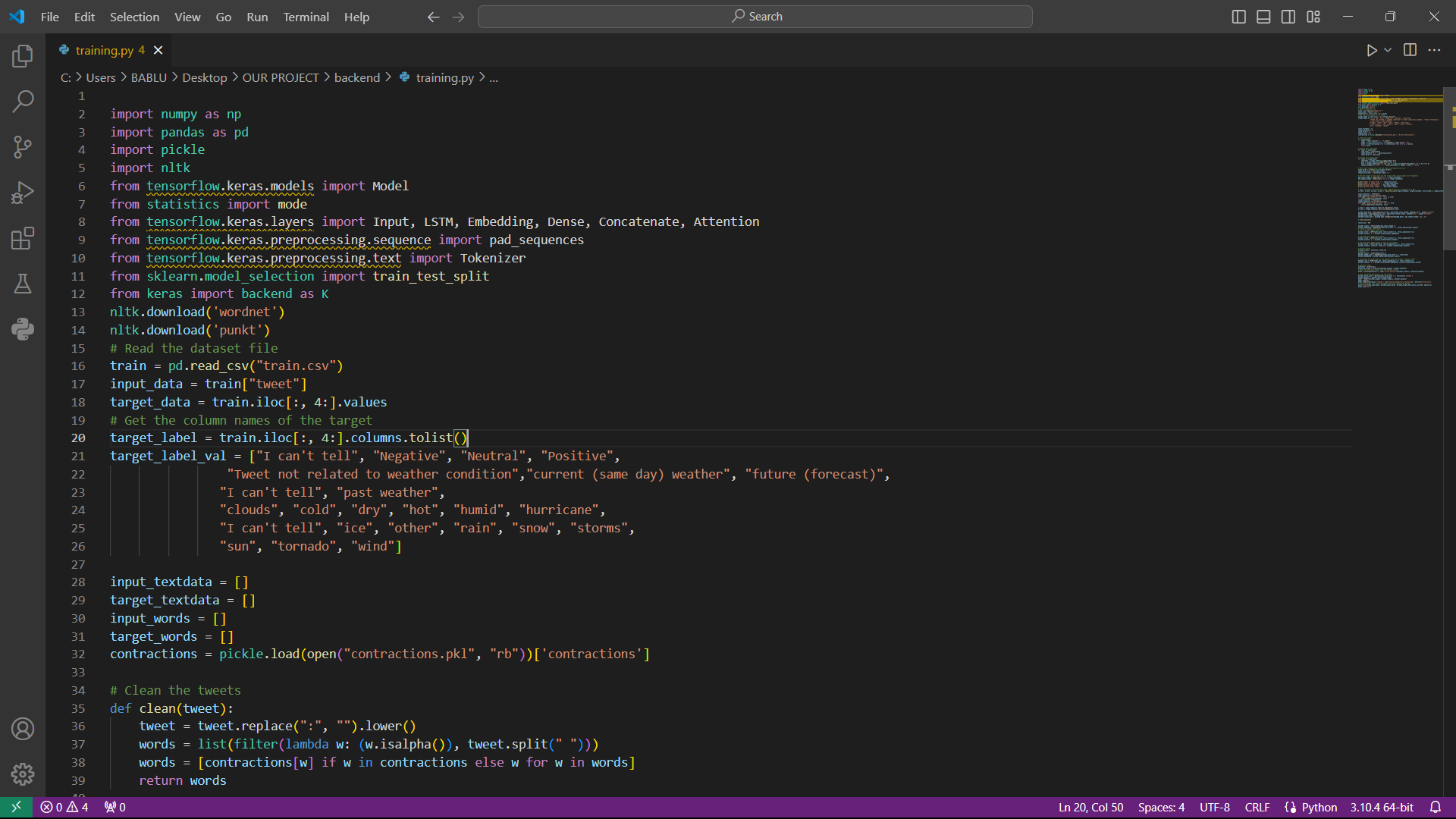
****

**Figure-4.2.2: Input and Preprocessing**

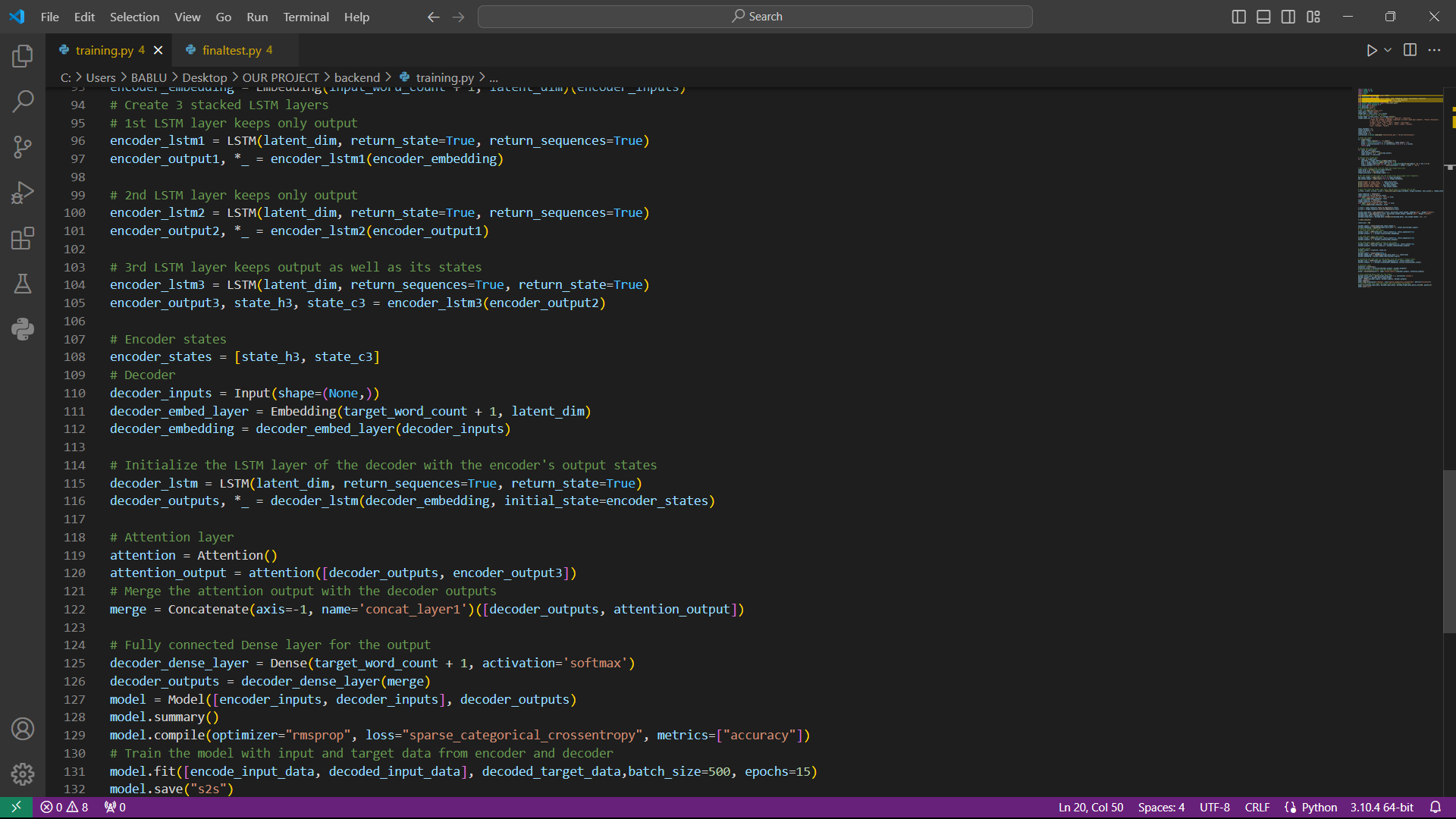


**Figure- 4.2.3: Testing**

# **5 IMPLEMENTAION, EXPERIMENTAL RESULTS & TESTCASES IMPLEMENTATION**



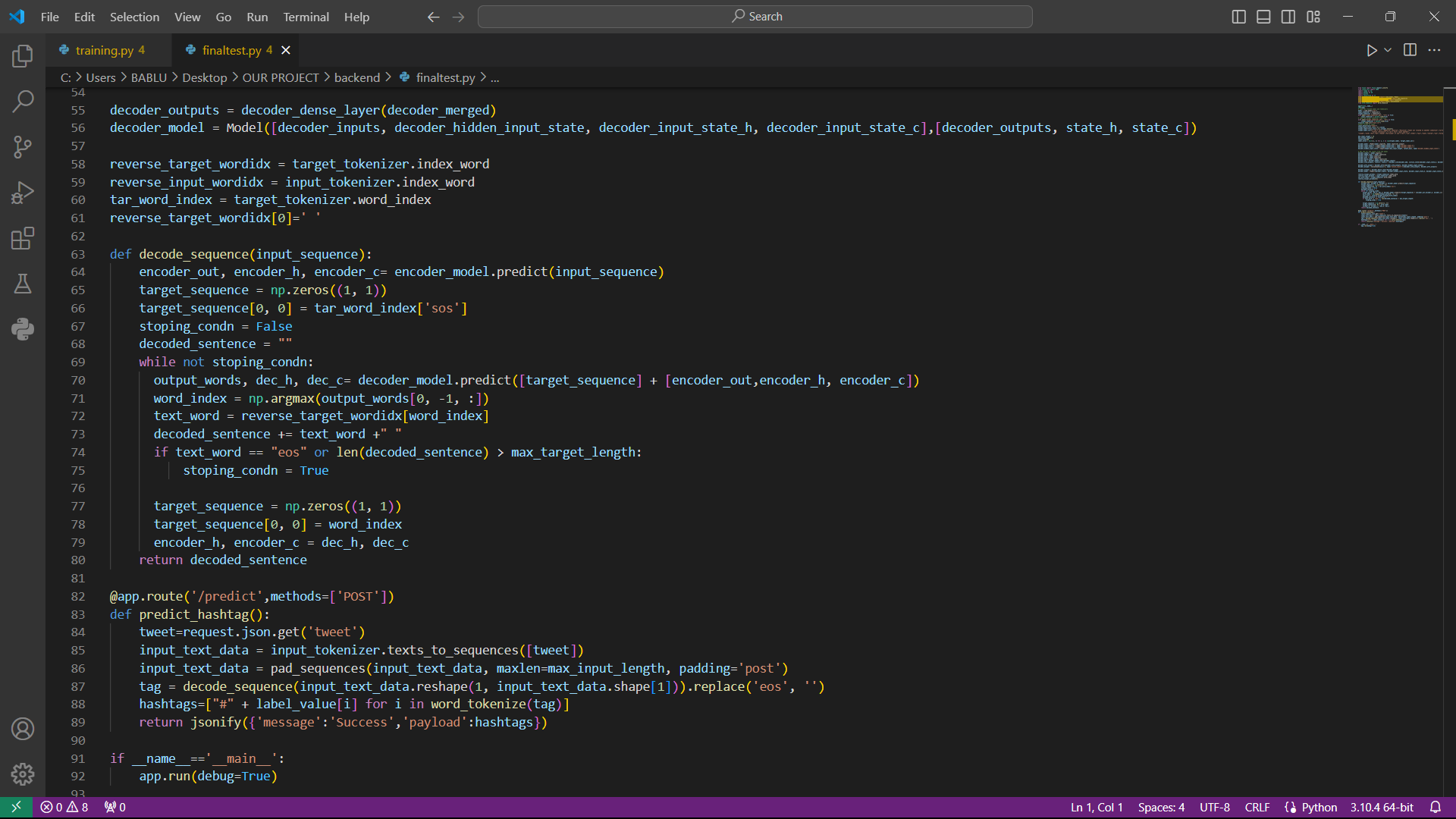
**Figure-2.1: Training Model**

****

**Figure-2.2: Training Model**

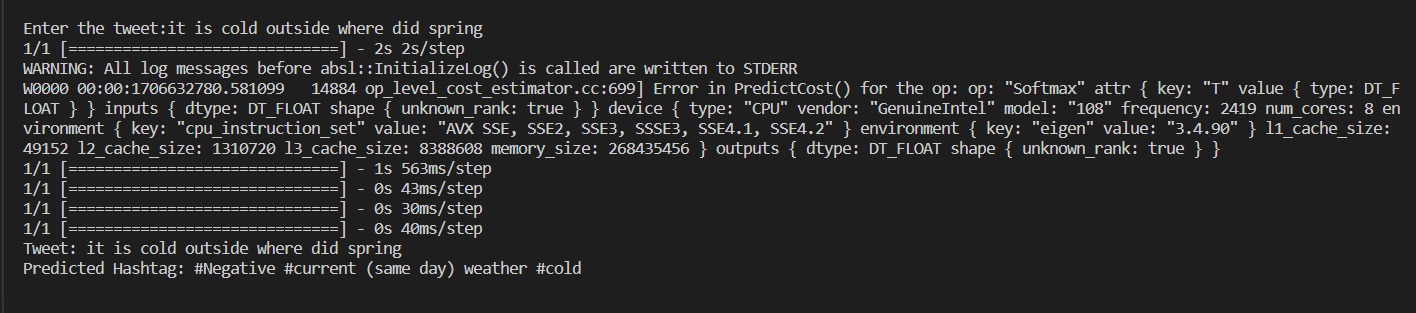
# 

**Figure-2.3: Testing**

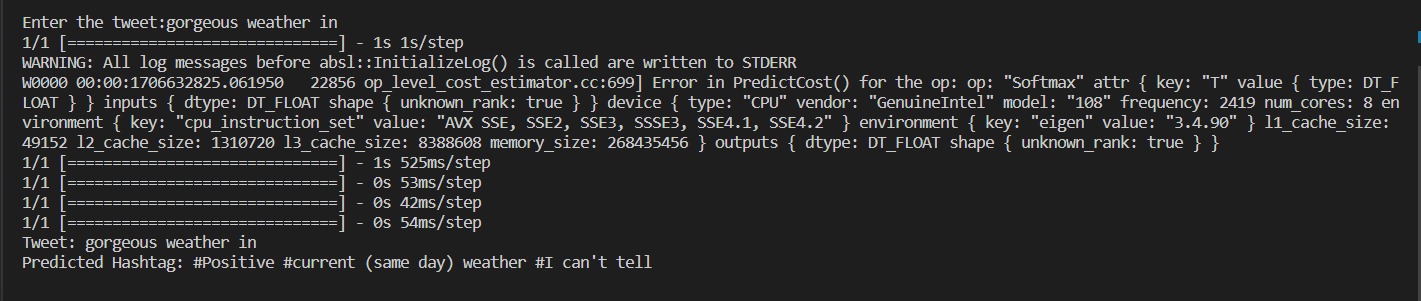
****

**Figure-2.4: Testing**

# **EXPERIMENTAL RESULTS**



**Figure-2.5: Results**

****

**Figure-2.6: results**

**TEST CASES**

|  |  |  |
| --- | --- | --- |
| **TEST CASE** | **TEST CASE SCENARIO** | **EXPECTED OUTCOME** |
| 1 | Test Case with one hashtag | The system identifies and return the single hashtag “#happy” |
| 2 | Test Case with multiple hashtags | The system identifies and return the multiple hashtags “#healthyliving” |
| 3 | Test Case with no hashtags | The system identifies and return an empty list as there are no hashtags present in the given input |
| 4 | Test Case with hashtags that contains special characters | It identifies and return hashtags contains special character “#sneak&peak” |
| 5 | Test Case with emojis as hashtags | It identifies correctly and return emojis used as hashtags like “#vacation” |
| 6 | Test Case with hashtags in different languages | It identifies and return hashtags in different languages like “#BastilleDay” and “#FeteNationale” |
| 7 | Test Case with multiple hashtags in one word | The system identifies and return each individual hashtag such as “#AI” , “#ML” and “#Tech” |
| 8 | Test Case with duplicate hashtags | The system should identifies and return only unique hashtags, eliminating duplicates |

# **6 .CONCLUSION AND FUTURE SCOPE**

**CONCLUSION**

An inventive way to improve user experience and content visibility on well-known social media sites is the Hashtag Prediction System. The technology offers real-time and customized hashtag recommendations by integrating smoothly with the content generation process through sophisticated Natural Language Processing and machine learning algorithms. The model is kept up to date with changing user patterns and behaviours thanks to the methods for continual learning and adaption. The technology involves users in improving the underlying prediction model in addition to providing intelligent hashtag predictions by allowing customization choices and incorporating user comments. In general, the initiative aims to transform hashtag usage and promote a more dynamic and user-centred approach to content production in the constantly changing social media ecosystem. The core strength of the system lies in its real-time hashtag prediction capabilities, providing users with intelligent and contextually relevant suggestions as they craft their social media posts. The underlying machine learning model, trained on diverse datasets, ensures adaptability to emerging trends and user preferences.

**SCOPE**

Future development of the social media hashtag prediction system looks promising, with the main focus being on improving user engagement and content production experiences. The project's course entails ongoing improvement of its forecasting abilities to provide users with a customizable and user-friendly hashtag recommendation engine. A more inclusive global reach is ensured by enhancing the system's adaptability to a wider range of languages and cultural contexts, meeting the individual preferences of users everywhere. Moreover, using cutting-edge technology like augmented reality and multimodal analysis has the potential to improve user experience and open up new avenues for creative content production. Furthermore, the incorporation of marketing tools can promote strategic partnerships with companies and influencers, which can further develop the system into an invaluable tool for content development and audience engagement tactics. Keeping these in balance

# **7 .REFERENCES**

* Bhat, S. A., Shah, M. A., & Bhat, G. M. (2022). A Comprehensive Survey on Hashtag Prediction Techniques in Social-Media. (Search for similar recent survey papers for a broad overview.)
* Yang, J., Counts, S., & Hoff, A. (2021). Hashtag Popularity Prediction on Twitter: A Benchmarking Study.
* Wang, C., Jiang, D., Liu, S., & Lyu, M. R. (2020). Hierarchical Hashtag Prediction with Multi-Modal Attention Network.
* Kim, J., Lee, J., & Kim, H. (2021). Predicting Social Media Hashtag Trends with Transformer-based Models.
* Cheng, Y., Shen, J., & Ma, C. (2022). Temporal Attention Enhanced Model for Hashtag Prediction in Social Media Texts.
* Gupta, P., & Koolagudi, S. G. (2021). Hashtag Recommendation for Instagram: A Deep Learning Approach.
* Nguyen, T. H., Nguyen, G. T., & Nguyen, T. T. (2020). Enhancing Hashtag Prediction in Social Media Texts Using Transformer-based Models.

# **APPENDIX**

# **Full Paper-publication Proof**

# **Snapshot of the Result**

