IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13352

Revolutionizing Sentence Completion: Pioneering a Machine Learning Paradigm for Next Word Prediction

Allamsetty Krishna Teja[1], Bathula Srinivasa Reddy[2], Talagadadeevi Jyothish Sai[3], Padala Bharath vamsi[4], Lavanya Kongala[5]

Computer Science & Engineering (of Affiliation), Koneru Lakshmaiah University (of Affiliation) Vaddeswarm, India[1-5]

Abstract: Writing lengthy lines is a bit tedious, but the text prediction feature on the keyboard makes things easier. In the field of the study of natural languages, Next-Word Prediction (NWP), Often referred to as language modelling,is a machine learning tool that has the ability to anticipate the word that will come after a letter in a phrase or sentence. Users may select a word at will from the list of suggested words provided by the system and it offers many different word substitutions. The long short-term memory (LSTM)formula can recognize previous text and predict words, which can be helpful to users in sentence construction. This approach uses letter-by-letter prediction, meaning it predicts a word when a letter forms a word.

Keywords: Machine Learning, Next-Word Prediction(NWP), LSTM,.NLP

I. INTRODUCTION

Today, many people, including children and the elderly, use computer programs to interact via text messages, which can still be difficult for them. An illustration of the word prediction problem is this. The next word related to the current word is suggested by the word prediction model. Reducing the number of keystrokes is the goal of prediction tools from. The following word prediction systems reduce keystrokes and misspellings, save time, and let you type entire words with a single keyboard.

In this field, natural language processing, or NLP, has become popular as away to improve text comprehension. It is common to analyze linguistic data and develop a set of attributes. However, the n-gram model has disadvantages, such as increased success rates due to the use of common hazard tables, which place unreasonably high demands on computational power. Naive Bayes was developed as a solution to this problem but it has the disadvantage of losing information due to the independence assumption.

Latent semantic analysis (LSA) was created to solve these problems. By studying word connections at different text levels, LSA evaluates text information semantically and makes real-time predictions. Models are used in machine learning to train and understand data. The method that Machine Learning applies is different from that of traditional programming. Conventional programming produces results in response to inputs such as data and rules. However, with machine learning, responses and data are used as input and rules are generated as output.

Furthermore, a model capable of identifying different types of models is built. Neural networks are used in Deep Learning. They make classification and prediction decisions Machine learning is closely related to next word prediction in the context of natural language processing (NLP). Next word prediction involves using algorithms and models that can learn patterns, relationships, and contextual information from text data to predict the most probable word that would follow a given sequence of words.

Correlation between Machine Learning and Next Word Prediction:

Pattern Recognition: Machine learning algorithms, especially those used in NLP, excel in recognizing patterns and relationships within textual data. They learn from large datasets, capturing statistical patterns in word sequences.

Feature Extraction: Machine learning models extract meaningful features from text data, representing words in numerical form (word embeddings). These representations capture semantic relationships between words, aiding in predicting the next word based on context.

Impact Factor 8.102 ⪽ Peer-reviewed & Refereed journal ⪽ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13352

Model Training: ML models, such as recurrent neural networks (RNNs), LSTMs, GRUs, and transformer-based models (like GPT, BERT), are trained on sequences of words to predict the next word. These models learn the underlying structure and context in text sequences, improving their ability to predict subsequent words accurately.

Contextual Understanding: Machine learning models leverage context-awareness, understanding the context of preceding words to predict the next word effectively. They learn semantic relationships, syntactic structures, and long-range dependencies within text data.

Continuous Learning: ML models can continuously improve and adapt based on new data. As they are exposed to more text examples, they update their internal representations, enhancing their predictive capabilities for next word suggestions.

In general, machine learning and next word prediction are related because they both use ML models and algorithms to understand, learn, and anticipate the most likely next word given the context given by words that have come before it in a phrase or sequence. These models improve user experience in a variety of text-based applications by producing precise predictions using learnt patterns and embeddings.

Overall, the correlation between machine learning and next word prediction lies in leveraging ML algorithms and models to comprehend, learn, and predict the most probable next word based on the context provided by preceding words in a sentence or sequence. These models use learned patterns and embeddings to generate accurate predictions, thereby enhancing user experience in various text-based applications.

Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) architecture that excels at capturing long-range dependencies and sequential patterns in data. LSTM networks have found significant applications in natural language processing tasks, including next word prediction.

LSTM Connection to Next Word Prediction:
Sequential Data Modeling: LSTMs are well-suited for modeling sequential data, such as language sequences. They maintain an internal state and effectively learn to capture dependencies between words in a sentence or text sequence.

Long-Term Dependencies: LSTMs address the vanishing/exploding gradient problems faced by traditional RNNs, allowing them to capture and remember long-term dependencies in text sequences. This capability is crucial for understanding context in next word prediction.

Memory Cells: LSTMs contain memory cells that store information over long sequences. These cells can retain relevant information while discarding unnecessary details, allowing the network to focus on important words and their contextual relationships.

Learning Temporal Relationships: LSTMs use gates (input, forget, output gates) to regulate the flow of information, enabling them to selectively remember or forget information over time. This mechanism aids in learning temporal relationships between words.

Context-Aware Predictions: LSTMs process sequences of words and, at each step, predict the probability distribution of the next word in the sequence. The network leverages the learned context from previous words to generate more accurate predictions for the next word.

Word Embeddings: LSTMs often use word embeddings (vector representations of words) as input. These embeddings capture semantic relationships between words and contribute to the LSTM's ability to understand and predict the next word based on contextual information.

Training and Optimization: LSTMs are trained using large text corpora, optimizing their weights and parameters through backpropagation and gradient descent. They aim to minimize prediction errors and maximize the likelihood of predicting the actual next word in the sequence.

In the context of next word prediction, LSTM networks excel in understanding the sequential nature of language, capturing dependencies, and leveraging learned context to predict the most probable next word given a sequence of preceding words. Their ability to retain long-term information and learn from extensive text data makes them a powerful tool in NLP tasks like language modeling and text generation.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13352

II. DESIGN/METHOD/MODELING

Next Word Prediction entails predicting the subsequent word that is highly probable to follow the current word. The system provides several word options, allowing users to select a word of their preference from a list of suggestions provided by the system. Next Word Prediction entails predicting the subsequent word that is highly probable to follow the current word. The system provides several word options, allowing users to select a word of their preference from a list of suggestions provided by the system.

The entire project is mainly divided into 4 modules namely

a. Dataset Creation Module
b. Data Processing Module
c. Model Selection Module
d. Model Training Module
e. Model Testing Module

a. Dataset Creation Module :

The foundation of any effective language modeling project lies in the quality and diversity of the dataset. In the context of our Next Word Prediction (NWP) project, the Dataset Creation Module plays a pivotal role in curating a dataset that captures the nuances of language usage.

1.Data Acquisition through Web Scraping:
The initial step involves acquiring textual data through a web scraping technique. This approach enables the collection of diverse language patterns from various online sources, ensuring a broad representation of linguistic styles. Web scraping is employed to fetch sentences and phrases from different channels, contributing to the creation of a comprehensive dataset.

2.Utilizing the Medium-Articles-Dataset:
One primary source for our dataset is the Medium-articles-dataset, selected for its richness in well-structured textual content. However, before integrating it into the training data, a thorough preprocessing step is executed to clean and organize the text. This ensures that the dataset aligns with the requirements of our language modeling task.

3.Tokenization and Sequence Generation:
Once the raw text is obtained, the Tokenizer is employed to convert the text into numerical sequences. Each word is assigned a unique numerical identifier, facilitating the model's understanding of the sequential nature of language. The module also generates input sequences by considering the relationship between words within sentences. This process involves creating sequences of increasing length, where each sequence represents a word's context within the sentence.

4. Ensuring Proper Padding:
To maintain consistency in input data dimensions, the sequences are padded to a fixed length. The padding ensures that all input sequences are of equal length, facilitating uniform processing by the model. This step is crucial for optimizing computational efficiency during training, as the model processes sequences in batches.

In summary, the Dataset Creation Module serves as the cornerstone for training our Next Word Prediction model. It involves the strategic acquisition of diverse textual data, careful selection of relevant datasets, and meticulous preprocessing to prepare the input data for effective language modeling. The resulting dataset is a vital component in training a robust and accurate model for predicting the next word in a given context.

b. Data Processing Module:

Efficient data processing is a critical component in the development of a robust Next Word Prediction (NWP) model. The Data Processing Module is responsible for transforming raw textual information into a format suitable for training the language model. This module encompasses various stages, from tokenization to the creation of input sequences, ensuring that the model can effectively learn the underlying patterns of language.

1.Tokenization and Vocabulary Construction:
The first step in data processing involves tokenization, where the textual data is broken down into individual words or tokens. The Tokenizer, a key component, assigns a unique numerical identifier to each word in the dataset. This process establishes a vocabulary that the model will use to understand and predict the next word based on the context provided.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13352

2. Sequence Generation and Padding:
Following tokenization, the module generates input sequences by considering the relationships between words within sentences. These sequences are incrementally constructed, with each step capturing an expanding context. To maintain uniformity, padding is applied to these sequences, ensuring that each input sequence has the same length. This step is essential for enabling batch processing during model training.

3. Data Splitting for Training and Validation:
To evaluate the model's performance accurately, the dataset is divided into training and validation sets. The Data Processing Module undertakes this task, allocating a portion of the dataset for training the model and another for validating its performance. This separation helps prevent overfitting and ensures the model generalizes well to unseen data.

4. Embedding and Preprocessing for LSTM Input:
The module incorporates an Embedding layer, translating the numerical word identifiers into dense vectors. This transformation enhances the model's ability to capture semantic relationships between words. Additionally, the data is preprocessed to create input sequences for the Long Short-Term Memory (LSTM) model, a key component in our NWP architecture.

In summary, the Data Processing Module serves as a crucial intermediary between raw textual data and the training of the NWP model. Through tokenization, sequence generation, and other preprocessing steps, this module prepares the data to be ingested by the model, facilitating the extraction of meaningful language patterns and enhancing the model's predictive capabilities.

c. Model Selection Module:

Selecting an appropriate model architecture is a pivotal decision in the development of a Next Word Prediction (NWP) system. The Model Selection Module involves carefully choosing and configuring the neural network architecture that best suits the nature of the language modeling task. In our project, two prominent recurrent neural network (RNN) architectures, namely Long Short-Term Memory (LSTM) and Bidirectional LSTM, were considered and rigorously tested to determine their efficacy in predicting the next word in a given context.

Name of the model	Family	Pretraining Architecture	Pretraining task(Method	Application	corpus
BERT	BERT	Encoder	MLM(MASKED LANGUAGE MODELING)&NSP (NEXT SCNTENCE PREDECTION)	comprehensive language comprehension	WebText &English corpus
BART	BERT	Encoder&Decoder	DAE	paraphrasing, language generation	WebText& English corpus
RoBERTa	BERT	Encoder	MLM(MASKED LANGUAGE MODELING)	named entity recognition, question answering	English corpus
XLNET	Transformer XL	Decoder	PLM(permutation Language Modeling)	text classification, language modeling, question answering, summarization	WIKPEDIS& English corpus
ELECTRA		Encoder	RTD(Replaced Token Detection)	text classification, sentiment analysis	WIKPEDIS& English corpus
XLM-RoBERTA	RoBERTa	Encoder	MLM(MASKED LANGUAGE MODELING)	machine translation, language understanding	News Articles English corpus

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13352

1.LSTM Model:
The Long Short-Term Memory (LSTM) model stands as a foundational choice for sequence modeling tasks. Leveraging the capability of LSTMs to capture long-term dependencies, this model is adept at understanding intricate patterns within language sequences. By employing an Embedding layer to map words into dense vectors and a subsequent LSTM layer for sequence learning, the model is trained to predict the next word in a sentence. The LSTM architecture excels in preserving contextual information over extended sequences, making it a suitable candidate for our NWP task.

2. Bidirectional LSTM Model:
Recognizing the importance of bidirectional context, the Model Selection Module explores the Bidirectional LSTM architecture. Unlike the unidirectional LSTM, the Bidirectional LSTM processes sequences in both forward and backward directions, enabling the model to capture context from both preceding and succeeding words. This bidirectional approach enhances the model's understanding of intricate language patterns, potentially improving its predictive accuracy. The Bidirectional LSTM is implemented with the same Embedding layer and LSTM layer setup, but with the added advantage of bidirectional information flow.

3. Comparative Analysis:
The Model Selection Module systematically evaluates and compares the performance of the LSTM and Bidirectional LSTM models. Both architectures undergo training for 50 epochs, and metrics such as accuracy and loss are meticulously assessed. This comparative analysis aims to discern the strengths and weaknesses of each model, providing insights into their predictive capabilities and aiding in the selection of the most effective architecture for our Next Word Prediction system.

In summary, the Model Selection Module navigates the intricate decision-making process of choosing the optimal neural network architecture. The consideration and comparison of LSTM and Bidirectional LSTM models contribute to informed choices in designing a language model that excels at predicting the next word in a given linguistic context.

d. Model Training Module:

Training the selected model architecture is a crucial phase in the development of a robust Next Word Prediction (NWP) system. The Model Training Module involves feeding the prepared dataset into the chosen neural network, fine-tuning the model's weights, and optimizing its ability to predict the next word based on contextual information.

1.Epochs and Training Time:
The training process involves iterating over the dataset for a set number of epochs. In our project, the models, including the Long Short-Term Memory (LSTM) and Bidirectional LSTM, are trained for 100 epochs. The number of epochs is a critical hyperparameter that determines the number of times the entire dataset is processed during training. This iterative approach allows the model to progressively learn and adjust its parameters to improve predictive accuracy. Monitoring the training time provides insights into the computational efficiency of the chosen architectures.

2. Loss Function and Optimizer:
The Model Training Module utilizes the categorical crossentropy loss function, suitable for multi-class classification tasks such as Next Word Prediction. This loss function quantifies the disparity between the predicted and actual word distributions. The optimizer, in this case, is Adam, a popular optimization algorithm known for its efficiency in adapting learning rates during training. The combination of categorical crossentropy and the Adam optimizer aims to guide the model towards more accurate predictions.

3. Model Evaluation and Validation:
To gauge the model's effectiveness, it is essential to evaluate its performance on data it has not seen during training. The dataset is divided into training and validation sets, with the validation set acting as a benchmark for generalization. The Model Training Module continually assesses metrics such as accuracy and loss on the validation set, providing valuable feedback on the model's ability to predict the next word across different contexts. This iterative evaluation process helps prevent overfitting, ensuring the model's proficiency on unseen data.

4. Learning Curve and Optimization:
Monitoring the learning curve, a graphical representation of the model's performance over epochs, aids in understanding its training dynamics. By observing the convergence of training and validation metrics, adjustments can be made to optimize the model further. Fine-tuning hyperparameters or considering alternative architectures may be explored based on the insights gained from the learning curve, contributing to an improved and more accurate NWP system.

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13352

In summary, the Model Training Module is a pivotal stage where the selected architecture is iteratively exposed to the dataset, refining its parameters to enhance predictive accuracy. The careful selection of epochs, choice of loss function and optimizer, continuous evaluation, and monitoring of the learning curve collectively contribute to the development of a high-performing Next Word Prediction model.

e. Model Testing Module:

The Model Testing Module is a critical phase in assessing the performance and generalization capability of the trained Next Word Prediction (NWP) model. After the model has undergone rigorous training on the dataset, this module involves evaluating its predictions on unseen data, simulating real-world scenarios where the system encounters novel language patterns.

1.Testing with User-Defined Input:
The Model Testing Module begins by allowing users to input custom phrases or sentences for the model to predict the next word. This interactive testing approach simulates real user interactions and assesses how well the model adapts to varying inputs. The system processes user-defined input, tokenizes it, and utilizes the trained model to predict the next word, providing immediate feedback to users.

2. Quantitative Metrics Evaluation:
Beyond user interactions, quantitative metrics are crucial for a comprehensive assessment of the model's performance. Metrics such as accuracy, precision, recall, and F1-score are computed on a separate test dataset. This dataset contains examples that the model has not encountered during training, ensuring an unbiased evaluation. These metrics quantify the model's ability to make correct predictions and its robustness in handling diverse language patterns.

3. Cross-Validation for Robustness:
To enhance the reliability of the Model Testing Module, cross-validation is employed. This technique involves splitting the dataset into multiple folds, training the model on different subsets, and testing it on the remaining data. Cross-validation helps assess how well the model generalizes to diverse linguistic patterns and minimizes the impact of data variability on the evaluation results.

4. Visualization of Prediction Results:
To provide a more intuitive understanding of the model's performance, the Model Testing Module incorporates visualization tools. Visual representations, such as confusion matrices or prediction heatmaps, offer insights into the model's strengths and potential areas for improvement. Visualization aids in identifying patterns and anomalies in the model's predictions, contributing to a more informed analysis of its behavior.

In summary, the Model Testing Module is integral to ensuring the effectiveness and reliability of the Next Word Prediction model. By combining user interactions, quantitative metrics, cross-validation, and visualization, this module provides a holistic evaluation, offering insights into the model's real-world applicability, generalization capabilities, and areas for potential refinement.

III. LITERATURE REVIEW

An examination of current systems as part of a literature review highlighted the value of word prediction models in facilitating communication, especially for slow writers.

We investigated machine learning techniques and showed that they are unable to generate correct grammars.

The development of Residual-Connected Minimum Gated Unit (MGU) and Multi-Window Convolution (MRNN) algorithms has shown improvements in reducing training time without sacrificing accuracy.

In particular, the bidirectional LSTM model demonstrated its effectiveness in managing large datasets and improved the accuracy of word prediction.[10]. In this study, we introduce LSTM and BiLSTM models for predicting the next word in Hindi, thereby improving users' typing speed and minimizing keystrokes.We have shown that the proposed His LSTM and BiLSTM models perform better than the current model for next word prediction in Hindi.[1]. An application of natural language processing that involves text mining is called "next word prediction." The proposed system can predict at least five phrases by comparing and using multiple techniques and determine the optimal algorithm for accurate outcome prediction.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13352

The proposed method was eventually refined to predict more than 10 words and compare multiple algorithms to find the one that best fits the data.[2]. One area of NLP, next word prediction, has its own issues with text mining. Pre-processing was therefore carried out inside the model to increase the model's forecast. The outcomes revealedthat when the model got closer to forecasting the each subsequent word, the loss significantly decreased with the quantity of tries while maintaining accuracy enhanced. In the future, efforts to enhance performance

The methods and approaches listed below can be used in a way that increases model complexity, employing word processing and regularisation methods methods for embedding, enhancing data sets, and Using optimisation for hyperparameters.[3]. Next word prediction is the art of speculating on the potential word that will come next in a phrase. There are several methods to guess the following word. The main goals of word prediction are to minimise key pressure, speed up typing, and cut down on text message composition time. Next word prediction speeds up communication as a consequence. Numerous techniques have been proposed to improve the performance level of the word prediction system.This survey was aimed at analysing different next word prediction systems. The three main categories of next word prediction techniques are hybrid, deep learning, and statistical techniques. The following word prediction techniques' efficacy was assessed. GRU was determined to have the most techniques while comparing different approaches for different languages.One area of NLP, next word prediction, has its own problems with text mining.Therefore, preprocessing was performed within the model to increase the model's predictions.The results showed that the accuracy remained increasing as the model got closer topredicting each subsequent word, while the loss decreased significantly with the number of trials.Future efforts to improve performance

The methods and approaches listed below include increasing model complexity, using text processing and regularization methods, embedding datasets, improvement methods, and optimization using hyperparameters.Can be used in any way.[4]. It is said that this work on Telugu next word prediction using LSTM is an attempt to investigate and further electronic communication time management. Using this approach in electronic communication—email, social media, etc.—will help users save time and accomplish tasks more easily. The test accuracy of 95.43 and the loss of 0.1799, which is a very excellent indication of perfection, can be observed from the findings. Observing the results with datasets much bigger than those already in use will be intriguing. In our next course of action, we want to create larger datasets than the ones we currently utilise, apply our models, and assess the results. As a result, our subsequent word prediction model on the given dataset performs rather well. The prediction's overall quality is strong. To enhance the model's prediction, several pre-processing actions and modifications can be applied. While our work primarily focuses on Telugu, other regional dialects can also be used with our LSTM model. [5] GRU based RNN has the ability to anticipate the next most relevant and acceptable Bangla word (or words) and phrase. demonstrated a noteworthy contribution to our study. For To explain the importance of employing RNN based on GRU, we have contrasted our suggested approach with alternative techniques that were employed by researchers for Bangla and other languages, and Among them, accuracy improved . While Unigram's accuracy for our suggested task is only 32.17%, for higher-order series like 5-, 4-, and Tri-gram gramme, there is a high accuracy rate (95.84%, 99.24% as well as 99.70%. Once more, the general correctness of this strategy would be more striking if we could make advantage of a bigger dataset that we have previously utilised for this project. Employing Bangla the corpus of data wasSince there is no ready-made dataset for Bangla language, we had to collect the dataset from various sources, which was a challenge.

In the future, we will try to collect a large dataset to improve the performance of the GRU based RNN for next word and sentence prediction in Bangla.Moreover, this study will serve as a tool for sustainable technology in industry, as its applications are diverse and can be used in different fields[7]. This research offers the MCNN-ReMGU model, which is built on a residual-connected MGU network and multi-window convolution combined with data regularisation technology, for the natural language word prediction challenge. We confirm the efficacy of residual-connected MGU network and multi-window convolution in obtaining high-dimensional features between locally nearby words and feature information between word sequences, respectively, using the PTB dataset. Simultaneously, investigations reveal that the residual link to the MGU network completely learns the long dependency relationship between word sequences and solves the vanishing gradient problem and network degradation. Additionally, batch normalisation and the L2-norm are used to successfully reduce over-fitting.

Overall experimental findings demonstrate a considerable performance improvement with the suggested CNN-ReMGU.[8]. This endeavour demonstrates that, despite Sinhala's high degree of inflectional complexity, word predictors are capable of anticipating the subsequent word in a word series. Recurrent neural networks have the qualities that allow them to be utilised to build highly accurate models that produce accurate predictions. Even though this study was limited to a specific domain, it might be expanded to include the entire Sinhala language with more resources. The transliterated dataset's findings demonstrated that the inclusion of Sinhala text in the dataset had no effect on the trained model's accuracy. As a result, more study on these kinds of models might be trained and developed for use in text editors for profit without requiring[9]

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⧓ Peer-reviewed & Refereed journal ⧓ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13352

IV. RESULT

```
1/1 [==============================] - 0s 450ms/step
what are the
1/1 [==============================] - 0s 30ms/step
what are the available
1/1 [==============================] - 0s 25ms/step
what are the available color
1/1 [==============================] - 0s 41ms/step
what are the available color options
1/1 [==============================] - 0s 24ms/step
what are the available color options for
1/1 [==============================] - 0s 24ms/step
what are the available color options for mercedes
1/1 [==============================] - 0s 24ms/step
what are the available color options for mercedes benz
1/1 [==============================] - 0s 28ms/step
what are the available color options for mercedes benz gls
1/1 [==============================] - 0s 31ms/step
what are the available color options for mercedes benz gls series
1/1 [==============================] - 0s 39ms/step
what are the available color options for mercedes benz gls series suvs
```

The Next Word Prediction (NWP) project successfully implemented and trained LSTM model for language modeling.

Through rigorous evaluation and testing, the models demonstrated commendable accuracy in predicting the next word, showcasing their proficiency in capturing intricate language patterns. User interactions with the system's predictive capabilities proved intuitive and effective, emphasizing the practical usability of the developed NWP model.

V. CONCLUSION

In conclusion, our NWP system represents a significant stride in the realm of language modeling, with a focus on user-friendly interactions and accurate predictions. The journey from dataset creation to model selection, training, and testing has equipped us with valuable insights into the dynamics of language patterns.

As we look forward, the lessons learned and the success achieved pave the way for future enhancements and applications of our Next Word Prediction model in diverse linguistic contexts. The project underscores the potential of machine learning to revolutionize language understanding, opening avenues for improved human-computer interactions and natural language processing.

REFERENCES

[1]. A. Tiwari, N. Sengar and V. Yadav, "Next Word Prediction Using Deep Learning," 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), New Delhi, India, 2022, pp. 1-6, doi: 10.1109/GlobConPT57482.2022.9938153.

[2]. P. Chilukuri, K. N. K. Reddy, K. N. V. S. Divyanjali, L. J. Royal, M. H. Chandana and D. M. Krishna, "A Novel Model for Prediction of Next Word using Machine Learning," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 393-400, doi: 10.1109/ICICCS56967.2023.10142737.

[3]. V. Rathee and S. Yede, "A Machine Learning Approach to predict the Next Word in a Statement," 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2023, pp. 1604-1607, doi: 10.1109/ICESC57686.2023.10193001.

[4]. S. T and S. S, "Survey On Next Word Prediction Techniques In Natural Languages," 2023 International Conference on Innovations in Engineering and Technology (ICIET), Muvattupuzha, India, 2023, pp. 1-6, doi: 10.1109/ICIET57285.2023.10220846.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13352

[5]. V. S. R, R. K. Nanduri, B. Pinni, M. M V and S. S. M, "Next Word Prediction in Telugu using RNN Mechanism," 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2022, pp. 98-104, doi: 10.1109/ICAISS55157.2022.10010963.

[6]. A. Neelima and S. Mehrotra, "A Comprehensive Review on Word Embedding Techniques," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 538-543, doi: 10.1109/ICISCoIS56541.2023.10100347.

[7]. O. F. Rakib, S. Akter, M. A. Khan, A. K. Das and K. M. Habibullah, "Bangla Word Prediction and Sentence Completion Using GRU: An Extended Version of RNN on N-gram Language Model," 2019 International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2019, pp. 1-6, doi: 10.1109/STI47673.2019.9068063.

[8]. J. Yang, H. Wang and K. Guo, "Natural Language Word Prediction Model Based on Multi-Window Convolution and Residual Network," in IEEE Access, vol. 8, pp. 188036-188043, 2020, doi: 10.1109/ACCESS.2020.3031200.

[9]. N. T. K. Naulla and T. G. I. Fernando, "Predicting the Next Word of a Sinhala Word Series Using Recurrent Neural Networks," 2022 2nd International Conference on Advanced Research in Computing (ICARC), Belihuloya, Sri Lanka, 2022, pp. 13-18, doi: 10.1109/ICARC54489.2022.9754174.