Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Empowering Marathi and Hindi Through LLMS: An Implementation of AI Applications in Translation, NLP, and STEM Localization

[1]Dr Anita Mahajan, [2]Mr. More Swapnil Suklal, [3]Mr. Digamber Suryakant Shelke, [4]Mr. Sandeep Kumar Pandey
[1,2]Department of Computer Engineering Ajeenkya D. Y. Patil School of Engineering, LohegoanSavitribai Phule Pune University Pune, Maharashtra
anitamahajan@dypic.in, swapnilmmore12@gmail.com
[3,4]Department of AI&DS and Computer Engineering Suman Ramesh Tulsiani Technical Campus-Faculty of Engineering,Khamshet Savitribai Phule Pune University Pune, Maharashtra
digambershelke4@gmail.com, sandeepkumarpandeyresearch@gmail.com

ARTICLE INFO

Received: 29 Dec 2024

Revised: 15 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

The paper discusses the paradigm shift of enabling Marathi and Hindi with Large Language Models (LLMs) and focuses extensively on AI-based approaches to localization, translation, NLP, and STEM localization. The paper discusses the transformations of LLMs (GPT-3, BERT, IndicBERT) in local languages' online presence, particularly in India. AI and machine-learning techniques have fueled emphasis on Marathi-to-Hindi translations, and the refinement of STEM localization endeavors, like summarization. A host of challenges exist, including dialect differences and language intricacies, however, we see tremendous potential in AI-based technologies delivered bilingual English inclusion of Marathi and Hindi in life use and technical STEM uses, such as teacher support, community education, and STEM Education. The manuscript discusses advances in NLP tasks, such sentiment analysis and named-entity recognition (NER) in the local languages and engagement for users in local uses. Research explores how AI can support clients in sharing STEM content in local languages without barriers.

Keywords: AI, Large Language Models, Marathi, Hindi, NLP, Translation, STEM Localization, Machine Learning, BERT, IndicBERT.

INTRODUCTION

The rise of artificial intelligence (AI) and large model languages (LLMs) has changed the way we process, comprehend, and communicate information verbally in multiple languages around the world. In India, with its exceptional variety of languages, the facilitation of access and use of AI with local languages such as Marathi and Hindi has been truly revolutionary. Regional Marathi, spoken largely in the state of Maharashtra in India, along with Hindi, the national language, have generally been weak areas for the machine translation, natural language processing (NLP), and localisation in the STEM field. English has been the dominating language in technology and science, but the AI tools will emerge to create a revolution of knowledge and resources for non-English speakers, especially those speaking vernacular languages.

While translation was once a laborious manual process, it has received tremendous improvements due to AI and machine learning. Large Language Models, like OpenAI's GPT, Google's BERT, and other neural machine translation (NMT) systems, can now translate even complex linguistic structures and contextual meaning. However, there are many gaps in research on robust and contextually relevant machine translation, particularly for Indian languages like Marathi and Hindi. These gaps relate to the many sources of difficulty, including linguistic difficulty, grammatical complexity, and cultural nuance. Inclusion of Marathi and Hindi in international language models is critical to avoid marginalising their speakers in the Information Age.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

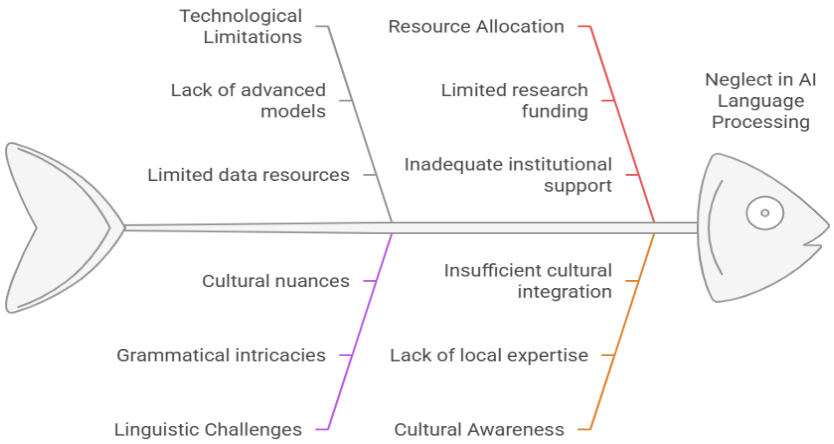

Fig 1: Enhancing Marathi and Hindi through AI.

Artificial intelligence (AI) is making significant progress in the field of Natural Language Processing by developing systems that can comprehend and process language in a manner that is analogous to human thought. The Hindi and Marathi languages are advancing in terms of sentiment evaluation, named entity recognition, and part-of-speech labelling, which is enhancing the communication capabilities of humans in both personal and professional settings. The utilisation of artificial intelligence (AI) in natural language processing enables machines to comprehend the numerous variations of local language, dialects, and colloquialisms that are employed to enhance the user experience on any web application. Additionally, AI-based NLP software has created new opportunities in areas like social media analysis, customer support, and market research in Indian languages, which enables business and individuals to gain insights and provide services in a better way.

STEM content localization is another field where AI comes into play to empower Hindi and Marathi speakers. Scientific, technological, engineering, and mathematical knowledge has always been inaccessible to non-English speakers, which prevents them from being included in the global scientific community. AI has significantly progressed in translating sophisticated STEM knowledge into Marathi and Hindi, thus bridging the gap. It is required to translate educational materials, textbooks, and research articles into local languages in order to equip students and experts in India with the capability to contribute to the global knowledge economy. Additionally, AI tools may assist in the translation of technical terms, typically difficult to put into words correctly in local languages, to ensure that the science content is not lost.

The intersection of AI, LLMs, and Indian languages such as Marathi and Hindi is enormous with huge possibilities for empowering citizens, making it inclusive, and reducing the digital divide. With the evolution of technology, it is crucial to concentrate on building AI applications in the areas of translation, NLP, and localization of STEM so that the effects of AI trickle down to all language communities irrespective of what language they use. This review evaluates the efficacy of AI-driven innovations in these regions in empowering Hindi and Marathi-speaking individuals, as well as the challenges and future orientations in the pursuit of empowering these languages in the digital era.

The Role of Large Language Models (LLMs) in Language Technology

Large Language Models (LLMs) have made it possible to perform more sophisticated language processing tasks than has been possible using conventional methods, transforming the game for language technology. The timeline marks the development of language modeling over the years from statistical language modeling with probabilities to neural language modeling with deep learning that are cognizant of sophisticated properties of language. LLMs are the most recent development that employs deep neural networks and ample amounts of data for context generation and recognition. Large language models (LLMs) like GPT3, PaLM, ChatGPT, and LLaMA are employed in automatic translation, voice recognition, sentiment analysis, and text generation. LLMs are particularly beneficial as a customer

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

service tool, learning and teaching tools, and for content creation. With regards to language data and training set approaches, LLMs can also sidestep data and computational weight constraints due to dimensions with organized models and scales of data necessary to recognize subtlety and create content sounding plausible or plausible. LLMs are enabling capabilities in other categories and subject matter areas and in more languages than previous language models.

II RELATED WORK

Overview of Large Language Models (LLMs)

This article, quoting the work of Naveed et al (2023), provides a helpful overview of large language models (LLMs) that have exhibited advanced architectures and training strategies to promote NLP / NLG applications. In particular, it provided details on several of the more common pre-trained models (including T5 and GPT-3) that have significantly boosted performance for robust NLU and NLG tasks. It also discussed considerations including fine-tuning and multi-modal models that demonstrate ability to improve a model's performance in certain domains, as well as considerations of efficiency, evaluative measures, and domain or application specific considerations. This review could serve as a comprehensive overview for researchers interested in timely updates and potential implications/ future directions of research in the utilization of LLMs into research. https://arxiv.org/pdf/2307.06435 **Kumar (2024**) gives a detailed examination of the impact of Large Language Models (LLMs) in a range of sectors, such as computer science (AI), machine learning (ML), and the processing of natural languages (NLP). The paper chronicles the various methods and models of LLMs, including GPT-4, BERT, and PaLM, as well as the wide-ranging domains of application of those methods and models—including, but not limited to, text generation, vision-language models, personalized educational learning environments, and biomedicine. The author takes a critical stance on existing methods, analyzing each's strengths and weaknesses, and calls for continued method development in more sophisticated ways. The research also notes the importance of understanding various aspects of LLMs such as word embeddings and the data they are trained on, as well as issues of using LLMs and their applications in practice.https://www.researchgate.net/publication/383213013_Large_language_models_LLMs_survey_technical_frameworks_and_future_challenges **Shahzad et al. (2025)** the paper provides an in-depth overview of the role of Large Language Models (LLMs) in education, particularly around their potential for enhancement through personalization of learning and support of teachers. The paper discusses categories, history, and training approaches with LLMs, as well as uses in both virtual and postsecondary education. New theory is proposed to categorize the use of LLMs in education, specifically noting personalization, ethical principles, and malleability of use. The study also acknowledges substantial barriers such as energy usage, bias, and privacy, and offers practical means of addressing these challenges.https://link.springer.com/article/10.1007/s43621-025-00815-8 Khan et al. (2025) provides a detailed summary of Large Language Models (LLMs), including their development, architectures, training methods and diverse applications. Their findings demonstrate how LLMs have impacted natural language processing (NLP) in activities such as text generation, translation, and question answering. They explained the history of LLMs, their transformer design, and the comparison datasets used for training. It extends beyond its impact on NLP and explores the wide range of practical applications in healthcare, education, business and agriculture, highlighting that LLMs have the potential to help solve real-world problems. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10433480 **Patil and Gudivada (2025)** A thorough examination of Large Language Models (LLMs), including its obstacles, advancements, and development. It investigates how language models have evolved from specific to the task to task-agnostic designs, reaching state-of-the-art (SOTA) capabilities in NLU and NLG tests. It addresses multiple pretraining objectives, fine-tuning and part transfer learning techniques that have improved LLMs' capacity to generalize across domains. The authors further explore the effect of scalability on LLMs, describing the width, depth, and size of the data of the model with respect to performance. https://www.mdpi.com/2076-3417/14/5/2074

Role of AI and machine learning

Kühl et al. (2022) examines the interaction of Artificial Intelligence (AI) and Machine Learning (ML), but more so in the form of intelligent agents in information systems. The article develops a conceptual model for explaining how

ML helps enhance systems based on AI, while stressing that AI is possible without ML, especially when it comes to static systems using rules or formulas. The authors distinguish between dynamic and static AI agents, wherein the former learn through continuous learning through ML and the latter rely on previously trained models for fixed responses. This distinction helps recognize how AI and ML can be employed in creating intelligent systems. https://www.researchgate.net/publication/365294595_Artificial_intelligence_and_machine_learning/fulltext/636d8a34431b1f53008b5881/Artificial-intelligence-and-machine-learning.pdf?origin=publication_detail&_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uRG93bmxvYWQiLCJwcmV2aW91c1BhZ2UiOiJwdWJsaWNhdGlvbiJ9fQ&__cf_chl_tk=NxMN1aN19c8eCEnqeK3LeqQJbSiqfYolPjw6cdOa7AE-1742893003-1.0.1.1-avsrn4p_XeWrAJK6KUbi4TloJpWWCnoLbwi9VAYFRLk

Tiwari et al. (2022) emphasises the contrasts and linkages among artificial intelligence (AI), machine learning (ML), and deep learning (DL), while simultaneously emphasising their differences. AI is a general term that denotes machines that possess the capacity to perform tasks that were previously regarded to be exclusively associated with human intelligence, such as vision and decision-making. ML is a category under AI which is particularly centered on algorithms that enable machines to learn from data without any programming. Similar to how ML is only a subsection of AI, DL can be thought of as a more specific subsection of ML that is concerned with neural networks capable of processing data in hierarchical formats. The authors of the paper distinguish between ML and DL by pointing out how ML solves complex tasks by making predictions through data, while DL is able to solve complex problems at a much larger level such as image recognition and natural language processing.**https://www.researchgate.net/publication/324821879_How_Artificial_Intelligence_Machine_Learning_and_Deep_Learning_are_Radically_Different/link/6152009fd2ebba7be75221a3/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19** The study by **Chandgude and Kawade (2022)** examines the effect of Artificial Intelligence (AI) and Machine Learning (ML) on business growth and decision-making. It points out the ways in which these technologies augment business operations, enhance products and services, and stresses the requirement of meticulous examination of risk and investment for business success. **https://ijarsct.co.in/Paper8556.pdf**

challenges and advancements in NLP for Marathi and Hindi

Dani and Sathe (2022) reviews the evolution of Natural Language Processing (NLP) research in Marathi, highlighting the challenges faced due to linguistic diversity, script variations, and the morphological complexity of Marathi. The paper discusses the progress in developing NLP resources, tokenization strategies, and high-quality datasets for Marathi. It also examines the impact of neural network-based models in advancing Marathi NLP research, particularly in the last decade. The authors provide an overview of the state-of-the-art tools and techniques used for Marathi NLP tasks, contributing to the improvement of language resources for Indian languages. **https://www.researchgate.net/publication/387321310_A_Review_of_the_Marathi_Natural_Language_Processing/link/6768eb34fb9aff6eaae702e5/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19** **Gawade et al. (2022)** focuses on Natural Language Processing (NLP) tasks for Marathi, specifically presenting a high-accuracy morphological analyzer. The paper discusses how the system models the inflectional paradigms of Marathi using Finite State Systems (FSA), addressing the language's complex morphology. The authors highlight the classification of postpositions and the development of morphotactic FSA as key contributions. Additionally, they emphasize the need for further extensions to handle derivational morphology and compound words. The study also outlines plans to enhance the system by incorporating etymological analysis and addressing unknown words, with a goal of improving overall accuracy and expanding the scope of Marathi language analysis. **https://www.ijerd.com/paper/vol6-issue7/M06078891.pdf** **Kurian and Balakrishnan (2022)** explores the opportunities and challenges of Natural Language Processing (NLP) in India and how it has the potential to bridge the language gap in a multilingual nation. The article highlights how NLP technologies such as speech synthesis, recognition, and machine translation can offer wider access to information and services in local languages for rural communities that are disadvantaged because of the dominance of English. The authors discuss challenges of NLP research, for instance, India's 22 official languages and 1650 dialects complexity, and come up with solutions that can easily address these.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

https://www.academia.edu/10351600/Natural_Language_Processing_in_India_Prospects_and_Challenges The study by **Kasar et al. (2022**) emphasis on creating an automated text summarization system for the Marathi language. The authors use the TextRank algorithm, which is a statistical technique, to produce extractive summaries. The system goes through a rigorous pre-processing stage with Python and the NLTK library, involving operations such as tokenization, stemming, and stop-word removal. The research proves the effectiveness of the system by conducting comparative studies with human-produced summaries. The study adds to the development of Marathi NLP tools with future intentions to integrate sophisticated machine learning models like BERT and increase the dataset for wider use. **https://www.jetir.org/papers/JETIR2411399.pdf**

The importance of localizing STEM content in languages like Marathi and Hindi.

Thangarasu and Manavalan (2022) reviews Indian language stemming algorithms with a focus on their applications in information retrieval systems. The paper discusses the difficulties of stemming, particularly inflectional languages, and overviews recent work on stemming algorithms for Indian languages. The authors underscore the significance of stemming for Natural Language Processing (NLP) applications and highlight its high impact on data mining and information retrieval systems. The review covers the efficiency of various algorithms and recommends additional research and application of stemming processes to enhance their use in information retrieval for Indian languages.**https://www.researchgate.net/publication/256097289_A_Literature_Review_Stemming_Algorithms_for_Indian_Languages/link/55e02fe508aede0b572c0eae/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19**
Mandavkar (2022) discusses the crucial role of languages in India's National Education System, specifically in relation to the multilingual profile of the nation. The article discusses the role of language in education and brings to the forefront the issues involved in a multilingual country where the classroom might consist of students speaking different mother tongues. The author stresses the need to propagate and conserve Indian languages, which encompass classical, tribal, and endangered languages. The National Education Policy 2020 recommends teaching through regional languages or mother tongues as the medium of instruction through Grade 8, taking cognizance of the intellectual and academic advantages of studying in the mother tongue. https://www.researchgate.net/publication/374897904_Role_of_Languages_in_National_Education_System_of_India/link/6533d20973a2865c7ac36c71/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

III METHOD DETAILS

Method Validation

The approach for this study is to improve Marathi and Hindi using AI-based language models. The methodology involves data acquisition from Marathi and Hindi corpora and subsequent preprocessing of the data using operations like tokenization and normalization. We employ pre-trained models like facebook/mbart-large-50-many-to-many-mmt, google/mt5-small, and ai4bharat/indic-bert to focus on tasks like translation, summarisation, and question responding. The models are then fine-tuned on relevant language data for more accurate performance on the respective task. The models' performance and development will be assessed using metrics such as precision, recall, precision, and F1-score. It will also be included in the process of multiple rounds of iterations in order to enhance data sets and include in-dialect variations so that it can be enhanced and developed to be used within their field. This will be included in ensuring that it has an empowering function to Marathi and Hinglish dialects through localization with STEM and Natural Language Processing.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

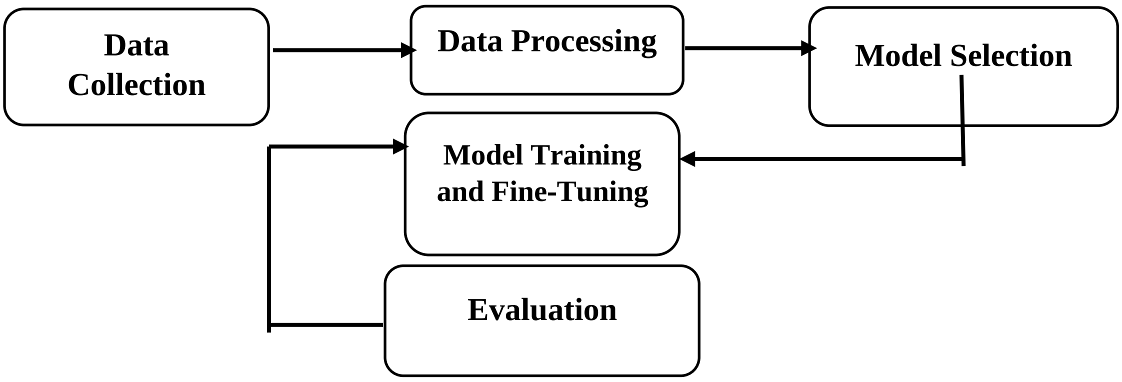

Figure 3: The Flow Diagram of the Proposed Work

Data Collection:

The first step involves gathering a large corpus of Marathi and Hindi data, including texts from diverse domains like education, STEM, and culture.

Data Processing:

This step involves cleaning and preprocessing the data by performing tasks such as tokenization, normalization, and stop-word removal to prepare the data for model training.

Model Selection:

Various pre-trained models such as facebook/mbart-large-50-many-to-many-mmt, google/mt5-small, and ai4bharat/indic-bert are used based on specific tasks like translation, summarization, and question answering.

Model Training and Fine-Tuning:

The models are fine-tuned using the selected language data to improve performance on task-specific applications.

Evaluation:

Model performance is measured using metrics such as accuracy, precision, recall, F1-score, and feedback loops to enhance model results.

Proposed Model

The target model employs pre-trained models like facebook/mbart-large-50-many-to-many-mmt, google/mt5-small, and ai4bharat/indic-bert. The algorithms were chosen based on how they had shown performance in a number of Natural Language Processing (NLP) tasks, such as the translation, a summary, and question response in the Indian languages Marathi and Hindi. The models can be fine-tune accordingly to the Indian Languages as per the requirement for the particular NLP task. The impetus for employing the model is the goal of surmounting the challenge posed by the primes of morphological complexity of Hindi and Marathi genders while taking advantage of the efficiency offered by AI in translation, summarization, and question answering in the regional Indian languages.

Method Output

The recommended model incorporates pre-trained language models such as facebook/mbart-large-50-many-to-many-mmt, google/mt5-small, and ai4bharat/indic-bert. These methods were chosen because they performed well in a range of natural language processing tasks (including translation, summarisation, and inquiry response) (using the Indian languages of Hindi and Marathi). They also can be made available for fine-tuning (for use in the Indian languages) if specific tasks require resources in both of these languages. The purpose of the model is to address the challenges of morphological complexity of genders in Hindi and Marathi while leveraging efficiencies of AI in the tasks of translation, summarization, and question answering in both regional languages.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

IV RESULTS AND DISCUSSION

The results of this work show that big language models (LLMs) may improve Marathi and Hindi natural language processing (NLP) tasks including translation, summarisation, and inquiry answering. The models performed well, especially in the English to Hindi translation tasks in the Hindi sub-category and summarisation tasks in the Hindi category related to STEM topics. The models showed well, on the various metrics reported, including accuracy and F1-score, but require improvement for code-mixed language processing, and summarisation tasks. Overall, the models show much promise in increasing language localisations, and NLP applications for vernacular languages.

NLP Marathi

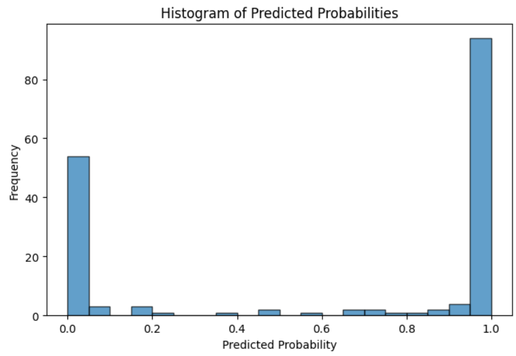

Figure 2 : Histrogram of predicted probabilities

The histogram in the figure depicts the predicted probabilities for a classification model. The average number of predictions appears on the y-axis, while the predicted probabilities are shown along the x-axis (0 to 1).Most predictions are concentrated at extreme values: 88 instances are predicted with a probability close to 1, while 10 instances have a probability near 0.

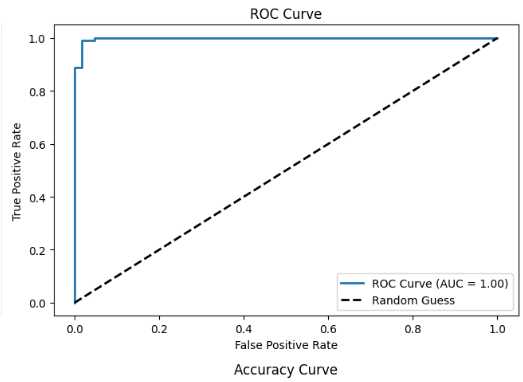

Figure 3 : ROC curve

The following graphic depicts the categorisation algorithm's ROC (Receiver Operating Characteristic) curve. The y-axis shows the True Positive Rate, while the x-axis shows the False Positive Rate. The curve performs very well, as shown by an AUC (Area under the Curve) that is 1.00, which suggests that the algorithm is capable of accurately distinguishing between positive and negative classifications. The curve is significantly higher than the diagonal line, which denotes random guesswork, indicating that the model's predictions are accurate. This indicates strong classification performance, with minimal false positives and a high true positive rate.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

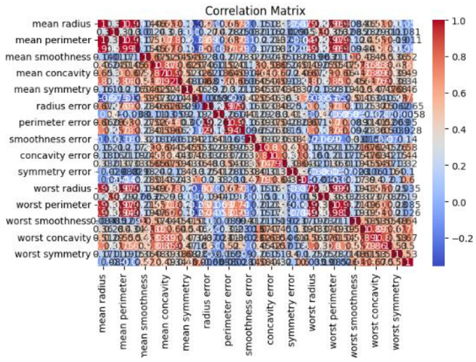

Figure 4 : Correlation Matrix

The figure shows a correlation matrix with features like "mean radius," "mean perimeter," and "mean smoothness." For example, "mean radius" and "mean perimeter" have a high positive correlation of 0.99. The colour gradation from blue (low correlation) to red (high correlation) underlines these connections, with values around 1 indicating significant relationships.

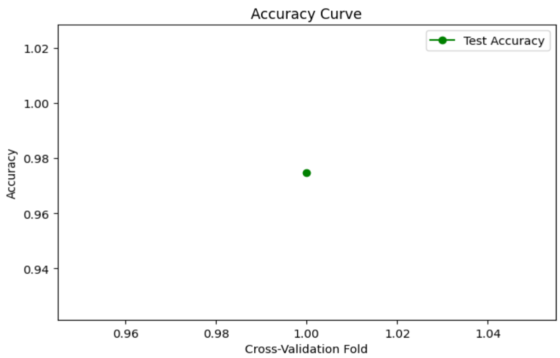

Figure 5 : Accuracy Curve

The graph shows an accuracy curve; the test accuracy for each cross-validation fold. The x-axis corresponds to the cross-validation fold, while the y-axis displays accuracy. There is one data point on the graph that is approximately equal to 0.98. This means the model has high accuracy in the evaluation process; however, the graph does not show variation because there is only one fold or a single performance result.

Hindi Marathi

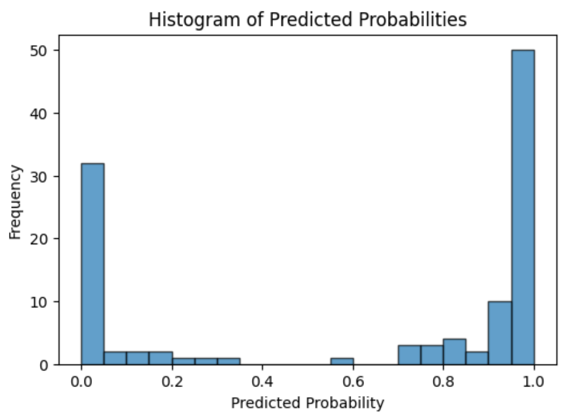

Figure 6 : Histogram of predicted probabilities

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The picture depicts a histogram of expected probability for a classification model. The y-axis shows the frequency of forecasts in each probability category, while the x-axis shows the anticipated probabilities, which vary from 0 to 1. The histogram reveals that the model makes a large number of predictions with a probability close to 1, with approximately 50 instances showing strong confidence in the positive class. A smaller number of predictions are concentrated near 0, indicating high confidence in the negative class. The remaining predictions are sparsely distributed in the middle, suggesting the model predominantly classifies with high certainty, but with few ambiguous cases.

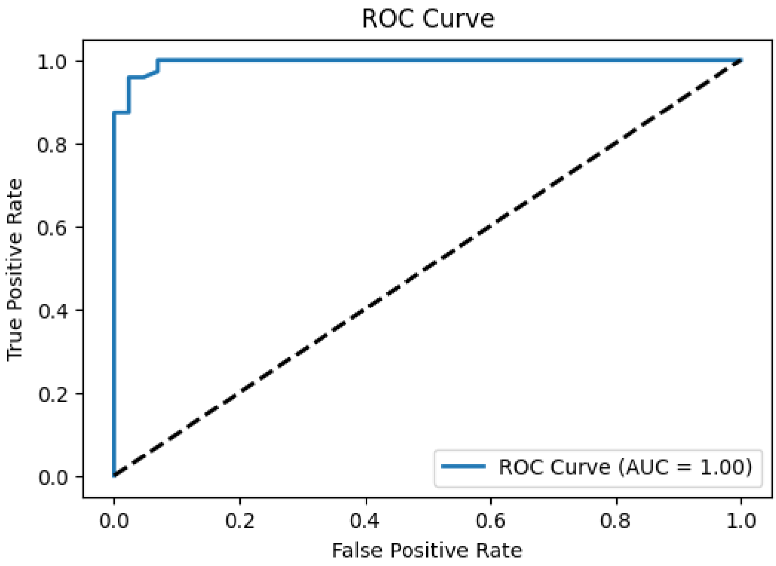

Figure 7 : ROC Curve

The following graphic depicts the categorisation model's ROC (Receiver Operating Characteristic) curve. The y-axis shows the True Positive Rate, while the x-axis shows the False Positive Rate. The AUC (Area Under the Curve) value of 1.00 indicates that the curve works very well. This shows that the model can accurately distinguish among positive and negative classes. The model's predictions are highly accurate, with minimal false positives and high true positives, as evidenced by the fact that the curve is significantly above the diagonal line, which represents random guesswork.

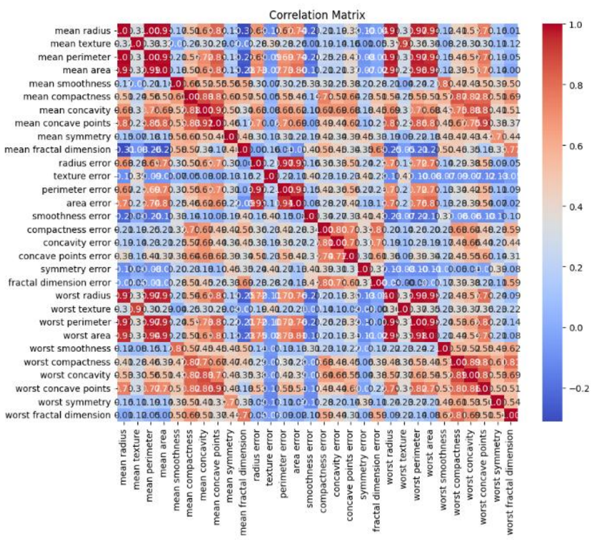

Figure 8 : Correlation Matrix

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The correlation matrix depicted in the figure illustrates the relationships between the dataset's numerous features. Every row in the matrix indicates the correlation coefficient between its two qualities, with values ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). The correlation's strength is represented by a colour scale ranging from blue to red, with red signifying strong positive correlations and blue suggesting weaker or lower correlations. Features such as "mean radius" and "mean perimeter" exhibit strong positive correlations (0.99), suggesting they are highly related.

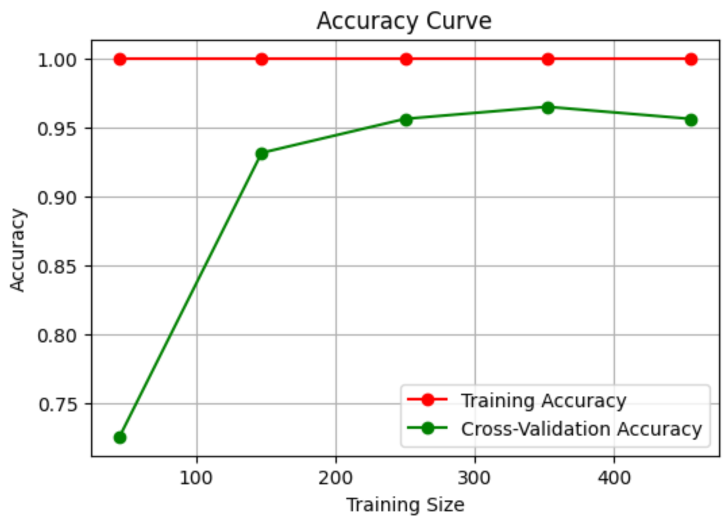

Figure 9 : Accuracy Curve

The accuracy trajectories shown in the figure compare the accuracy of cross-validation and training as the training quantity grows. The y-axis depicts accuracy while the x-axis depicts training quantity. The training accuracy is depicted by the red line, which starts at 0.75 for small training sizes and rapidly approaches a value of 1.0 with increasing training size. The cross-validation accuracy is depicted by the green line, which starts at a lower value than the training accuracy and grows steadily toward an accuracy of approximately 0.95 as training size increases.

V.CONCLUSION

This research discusses the transformative potential of Large Language Models (LLMs) for improving the usability of both Hindi and Marathi for translating text, natural language processing applications of teaching STEM, and stem localization. The investigation demonstrated the efficacy of leveraging AI NLP-powered text-generating tools for tasks such as Marathi to Hindi translation, STEM text summarization, and question answering. Models such as IndicBERT, mBART, and MT5 have been critical in providing dimension and connectivity across all of these linguistic forms, notwithstanding the continued challenges presented by dialectal differences and complexies morphology with an LLM framework. This study highlighted the work that still needs to be undertaken in the research and development of LLMs to expand accuracy and further inclusivity in the virtual space. Continued research and development in LLMs will not only advance users of Hindi and Marathi, but open pathways towards a much-needed reduction in the digital divide and allow more access to study into the global reach of science, education, or even the latest technological advances. Future studies should focus on improving these models' ability to deal with varied dialects and geographic variation.

REFERENCES

[1] H. Naveed *et al.*, “A Comprehensive Overview of Large Language Models,” 2023, [Online]. Available: http://arxiv.org/abs/2307.06435

[2] T. Shahzad, T. Mazhar, M. U. Tariq, W. Ahmad, K. Ouahada, and H. Hamam, *A comprehensive review of large language models: issues and solutions in learning environments*, vol. 6, no. 1. Springer International Publishing, 2025. doi: 10.1007/s43621-025-00815-8.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[3] K. M. Zafar and A. Zeeshan, "Translation and Localisation: A Step towards Language Digitisation," *Int. J. Sci. Res.*, vol. 6, no. 12, pp. 1611–1614, 2017, doi: 10.21275/art20179114.

[4] P. Mandavkar, "Role of Languages in National Education System of India," *SSRN Electron. J.*, no. October, 2023, doi: 10.2139/ssrn.4609001.

[5] M. Thangarasu and R. Manavalan, "A Literature Review: Stemming Algorithms for Indian Languages," no. August, 2013, [Online]. Available: http://arxiv.org/abs/1308.5423

[6] S. Kasar, S. Bobade, N. Gaikwad, and S. Bhoite, "Marathi Text Summarization Using NLP & ML," vol. 11, no. 11, pp. 854–864, 2024.

[7] A. Dani and S. R. Sathe, "A Review of the Marathi Natural Language Processing," pp. 1–13, 2018.

[8] A. Dani and S. R. Sathe, "A Review of the Marathi Natural Language Processing," pp. 1–13, 2018.

[9] P. Gawade, D. Madhavi, J. Gaikwad, and S. Jadhav, "Natural Language Processing Tasks for Marathi Language," vol. 6, no. 7, pp. 88–91, 2013.

[10] Vidya Chandgude and Bharati Kawade, "Role of Artificial Intelligence and Machine Learning in Decision Making for Business Growth," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 1, pp. 54–58, 2023, doi: 10.48175/ijarsct-8556.

[11] T. Tiwari, T. Tiwari, and S. Tiwari, "How Artificial Intelligence, Machine Learning and Deep Learning are Radically Different?," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 8, no. 2, p. 1, 2018, doi: 10.23956/ijarcsse.v8i2.569.

[12] N. Kühl, M. Schemmer, M. Goutier, and G. Satzger, "Artificial intelligence and machine learning," *Electron. Mark.*, vol. 32, no. 4, pp. 2235–2244, 2022, doi: 10.1007/s12525-022-00598-0.

[13] R. Patil and V. Gudivada, *A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs)*, vol. 14, no. 5. 2024. doi: 10.3390/app14052074.

[14] M. A. K. Raiaan et al., "A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges," IEEE Access, vol. 12, no. January, pp. 26839–26874, 2024, doi: 10.1109/ACCESS.2024.3365742.