



## A Hybrid Travel Recommender Model Based on Deep Level Autoencoder And Machine Learning Algorithms

Muneer V.K<sup>1\*</sup>, Mohamed Basheer K.P<sup>2</sup>

<sup>1\*, 2</sup>*Sullamussalam Science College, Affiliated to University of Calicut, Kerala  
vkmuneer@gmail.com<sup>1</sup>*

**\*Corresponding Author: Muneer V.K**

*\*Sullamussalam Science College, Affiliated to University of Calicut, Kerala  
vkmuneer@gmail.com*

<i>Article History</i>	<i>Abstract</i>
Received- 20 September 2023 Revised - 17 October 2023 Accepted: 22 November 2023	<p>This research investigates the application of autoencoders in processing travelogues written in the Malayalam language on Facebook. The main objective is to harness the capabilities of autoencoders to learn a compressed representation of the input data and employ it to train various machine learning models for enhanced accuracy and efficiency. The major challenge of unavailability of a benchmark dataset in the Malayalam language for the travel domain was overcome by employing NLP techniques on the unstructured, lengthy, imbalanced travelogues, applying some additional filtering methods, and the creation of an exclusive Part of Travel Tagger (POT Tagger) along with lookup dictionaries. As this pioneering work focuses on Malayalam travel reviews posted on social media, the model presents a valuable opportunity for extension to other low-resourced Indian languages. The study follows a two-step approach. Initially, an autoencoder neural network architecture is utilized to encode the travelogues into a lower-dimensional latent space representation. The encoder network adeptly captures crucial features and patterns within the data. The compressed representation obtained from the encoder is then fed into the decoder, which reconstructs the original travelogues. Subsequently, the encoded model is employed to train diverse machine learning models, including logistic regression, decision tree classifier, support vector machine (SVM), random forest classifier (RFC), K-nearest neighbours (KNN), stochastic gradient descent (SGD), and multilayer perceptron (MLP). By utilizing the encoded features as inputs, these models effectively learn from the concise representation of the Malayalam travelogues. Experimental results reveal that the trained machine learning models, using the encoded features, achieve higher accuracy rates compared to conventional approaches. This improvement demonstrates the effectiveness of autoencoders in capturing and representing vital characteristics of the Malayalam travelogues on Facebook. By leveraging capabilities of autoencoder model, we successfully learned a compressed representation of the input data, attaining an impressive validation accuracy of 95.84%. This finding highlights the potential of autoencoders to enhance the overall accuracy and efficiency of travel recommendation systems for Malayalam users on social media platforms.</p>

<b>CC License</b> CC-BY-NC-SA 4.0	<b>Keywords: Autoencoders, self-supervised learning, Recommender model, Natural language processing, Malayalam.</b>
--------------------------------------	---

## 1. Introduction

Travel and tourism play a pivotal role in promoting cultural exchange, economic growth, and personal enrichment. With the exponential growth of social media platforms and online travel communities, individuals now have unprecedented access to a wealth of travel-related information and experiences shared by fellow travellers. Among these platforms, Facebook travel groups have emerged as dynamic hubs for travellers to document their journeys, share captivating travelogues, and exchange valuable insights and recommendations. In this research paper, we focus on harnessing the vast potential of Facebook travel groups to develop a personalized travel recommender model. Our goal is to create a system that leverages the power of autoencoders and machine learning algorithms to provide tailored travel recommendations to users based on their individual preferences and interests. To achieve this, we collected a substantial corpus of 12,500 travelogues, each contain 40 sentences in average, written in the Malayalam language by scraping data from various Facebook travel groups.

Malayalam, the official language of Kerala, a southern state of India, and the Lakshadweep Islands, presents unique challenges in text and speech processing due to its intricate morphology, agglutinative nature, and rich inflectional structure. As a low-resourced language, it lacks uniformity in spelling and sentence structure across various sources within the State. Moreover, the unavailability of benchmark datasets poses a significant hurdle in developing robust models for text and speech processing tasks in Malayalam. These factors collectively contribute to the complexities faced by researchers and developers in harnessing the full potential of this unique language for natural language processing and speech-related applications. To ensure the quality and reliability of our dataset, we implemented rigorous data cleaning processes, resulting in a structured dataset with key features identified as Travel Type (TT), Travel Mode (TM), Location Type (LT), Location Climate (LC), Users (U) and specific destinations (L) mentioned in each travelogue. These crucial features form the foundation of our personalized travel recommender model.

The centrepiece of our approach lies in the application of autoencoder neural network architecture. By adopting this deep learning technique, we seek to study the reduced version of the travelogue data, effectively capturing essential patterns and features intrinsic to each travel experience. This encoded representation serves as a powerful tool to train various machine learning models, enabling them to learn from the compact and meaningful representation of the travel data. In this paper, we present our research findings, including the architecture of the autoencoder, the details of the data cleaning and preprocessing procedures, and the results obtained from training the machine learning models using the encoded features. We also highlight the validation accuracy achieved, which stands at an impressive 95.84%. The implications of our study extend to enhancing travel experiences by offering personalized travel recommendations that align with individual preferences, thereby promoting a more enriching and fulfilling travel journey.

Key Contributions are:

1. Autoencoder Model for Malayalam Travelogues: This research introduces the construction of an autoencoder model specifically designed for processing travelogues in the Malayalam language. The autoencoder serves as a powerful tool for unsupervised learning, capturing underlying patterns and features present in the travel data.
2. Analysis with and without Data Compression using autoencoder architecture.
3. Evaluation of Compressed Data: The research evaluates the performance of the autoencoder model when applied to compress travel data.
4. Enhanced Travel Recommendation Accuracy: By utilizing the encoded representations obtained from the autoencoder, machine learning models are trained for travel recommendation tasks.
5. Comparative Analysis of Machine Learning Models: The study includes a comparative analysis of various machine learning models, such as logistic regression, decision tree classifier, SVM, random forest, KNN, SGD, and MLP, all trained using the encoded travel representations.

## 2. Background and Related work

NLP advancements enable robots to read and analyse human language with remarkable precision, leading to a transformative impact on text understanding [1]. Mary Priya S conducted experimental investigations involving Statistical Machine Translation (SMT) for the Malayalam language and explores the impact of adapting SMT methods designed for foreign languages to a Dravidian language like Malayalam [2]. In Paper [3] R.K. Thandil explain the challenges on performing language processing on Malayalam language as its complex agglutinative nature and morphological richness.

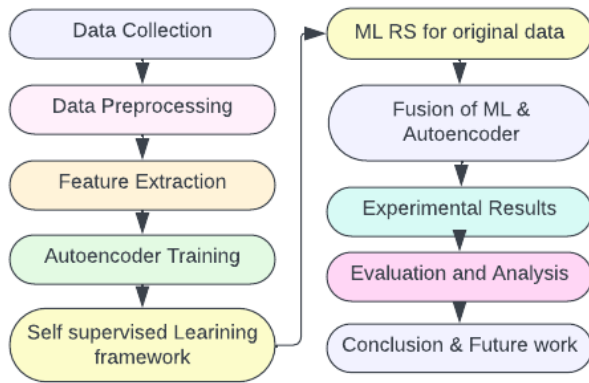
A systematic study on development of Recommender system done by D. Roy [4] demonstrated various strategies for information filtering, its types, their characteristics, and challenges. Personalized travel recommendation systems have garnered significant attention in recent years due to the increasing demand for tailored travel experiences [5]. F. Lu and W. Zhang in paper [6] proposed a new model for Personalized travel recommendation by combining the Graphical information on Travelogues (TGI) and points of interest (POI) algorithms.

Guijuan ZHANG and Yang LIU conducted a comparative study about autoencoder-based recommender systems and conventional recommender systems and concluded that autoencoder based RS outperforms the traditional models [7]. Ferreira D, et.al, proposed another product recommendation model that utilizes autoencoder-based on collaborative filtering, showcasing its superior performance compared to Singular Value Decomposition with a significantly low Root Mean Squared Error (RMSE) value [8]. In [9], Ding introduced an innovative privacy-preserving approach that adapts autoencoder for federated collaborative filtering, guaranteeing data privacy while achieving superior model performance.

Guo et al. [10] introduced an innovative hybrid recommendation system known as AutoLFA, that amalgamates methods of LFA and Autoencoder. AutoLFA utilizes two separate recommendation systems, each functioning within its distinct metric representation space, leveraging their respective advantages, and amalgamates them through a tailored self-adaptive weighting mechanism to harness the strengths of both methodologies. Y. Bougteb, et. al. [11], investigated a hybrid recommender system that combines a deep autoencoder for learning user interests and reconstructing missing ratings, along with SVD++ decomposition to capture correlations between different feature factors. In paper [12] Kamble studied the performance of recommender systems by leveraging data mining techniques by using an SVM-based recommender system and conducts experiments on various product datasets. A hybrid RS consists of deep learning model like autoencoders and various machine learning [13] models like SVM, KNN, logistic regression, decision tree, Random Forest and Multilayer perceptron are compared here in this paper.

## 3. Methodology

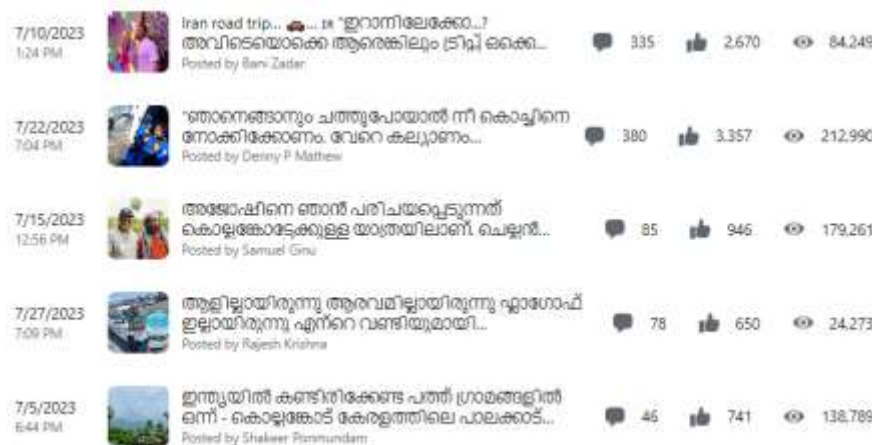
The methodology for developing a travel recommendation system using an autoencoder in this study is a multifaceted and cohesive process. Beginning with data collection, where information is sourced from community sites and online travel groups, the process continues with data preprocessing to transform the unstructured and noisy data into a consistent format. Feature extraction follows, where essential attributes are distilled from the text, facilitating dataset preparation. An autoencoder model is then constructed and trained to learn the compressed representation of travelogue data, capturing essential patterns and features. The self-supervised learning framework enhances the efficiency of the model, allowing for further refinement. Next, a machine learning recommender system is designed to work with both original data and the encoded representations obtained from the autoencoder, offering a fusion of traditional ML techniques with deep learning. Finally, experimental results are analyzed, and various models are compared, providing insights into the effectiveness and potential improvements of the travel recommendation system. The integration of these stages forms a comprehensive and robust methodology that tailors travel recommendations to individual preferences and experiences. The entire methodology is divided into the following phases as given in Figure 1.



**Fig. 1** Step by step methods of proposed work.

**4. Data collection**

In this study, our data collection process involved sourcing information from diverse community sites, with a primary focus on the largest Malayalam Travel group on Facebook, known as 'Sanchari,' as well as various travel blogs containing reviews and travelogues. Employing web scraping techniques, we gathered online write-ups contributed by random users. However, the process came with significant challenges, including client-side rendering, memory leaks, and bot detection mechanisms, which required careful navigation. The data obtained from these sources was highly unstructured, noisy, and inconsistent. Figure 2 shows the structure of list of posts(travelogues in Malayalam), its reactions, user name and profile picture, total count of viewed people and posted date and time. These posts are retrieved into an excel sheet which is the foundation of data collection and dataset preparation.



**Fig. 2** The list of travelogues and associated details in observed in Facebook travel group.

To maintain organization, we stored each write-up in separate individual files, named after the respective traveler, ensuring easy retrieval and management. Moreover, our system accommodates input from new users, capturing their individual travel-related text and storing it in separate files, facilitating seamless incorporation of new data. After extensive preprocessing, we transformed the collected 12,500 unstructured travelogues into a structured tabular format. A sample form of scraped travelogue is given as figure 3 which contain message as Travelogue, posted time, url of post and profile link, total reactions, likes, comments and its shares.

```

{
  "message": "ഇടപെട്ടിന്തിൽ നിന്നും നടവടയിലേക്കൊരു സൈക്കിൾ സ്വരവ് അതിവേഗം പോയി PLAF A ഹോളിങ്ങ് സ്വരവ് കൂടാതെ അതിൽ ഇരു കൈയപ്പലു അടെ കട്ടിൽ PLAF B ഉണ്ടാകൂവാ സെറ്റിംഗ് അക്ഷയം അന്നും പിന്നെ ഹെൽമെറ്റ് കൂടി വിട്ടിടാനും അലിയലളളതാണ്ട് വരാനു സഹായം വന്നു വിളിക്കണം താമസം ഒര അല്പ ദിവസം വരുന്നിടത്ത് പലതോളം 2.30 ടെ വീട്ടിൽ നിന്നും ഇറങ്ങി വലിയ പാതിയിൽ ഒന്നും ഇപ്പോൾ കോവുക തന്നെ അതിനുള്ള സൈക്കിൾ അതേ സ്വരവ് കൂടാതെ വളരെ സ്വരവ് കൂടി അങ്ങൽ ചവിട്ടി സ്വരവ് കൂടി വരും കഴിഞ്ഞപ്പോൾ ഒരു ഇറക്കുകിണം .....",
  "time": "2020-10-17T15:42:06.000Z",
  "post_url": "https://m.facebook.com/groups/758863974804742?view=permalink&id=3566357596755352",
  "posted_by": "Soufal Mani",
  "profile_url": "https://www.facebook.com/soufal.mani.526",
  "totalReactions": 1100,
  "comments": 100,
  "shares": 19,
  ".....":
}

```

**5. Travelogue preprocessing**

Text preprocessing is a crucial step in preparing the unstructured travelogues for the development of our personalized travel recommender model. The first step involves tokenization, where the text is divided into individual tokens or words, enabling us to work with discrete units of information. Following tokenization, the data undergoes cleaning to remove any irrelevant characters, symbols, or special characters, streamlining the text for further analysis. Additionally, we perform stopwords removal, eliminating common words that do not carry significant meaning, reducing noise in the dataset. Table 1 listed few among the stop words used in Malayalam language. To ensure linguistic consistency, we apply stemming or lemmatization, reducing words to their root or base forms, enabling us to group variations of a word together. This process further aids in text normalization. Moreover, we map certain words from a predefined dictionary, resolving synonyms and related terms to a standardized representation, enhancing the model's performance by reducing ambiguity. Incorporating domain-specific knowledge, we utilize a Part of Travelogue Tagger (POTT) to label and identify travel-related entities like locations, activities, and transportation modes. The tagged data is then organized and stored in CSV or Excel files for easy retrieval and analysis.

**Table 1.** Sample Stop words, Tokens, Lemmas in Malayalam, and corresponding English translation.

Stop words		Tokens and lemmas in Malayalam and English		
Malayalam	English	Tokens	Lemmas	English
അത്	that	പോയിരിക്കും	പോകുക	Go
അങ്ങനെ	so	കുടുംബമായി	കുടുംബം	Family
മതി	enough	കാറിലേക്ക്	കാർ	Car
എങ്കിൽ	if	തണുപ്പിന്റെ	തണുപ്പ്	Cool
മറ്റു	other	മരത്തിന്റെ	മരം	tree

**6. Feature Extraction**

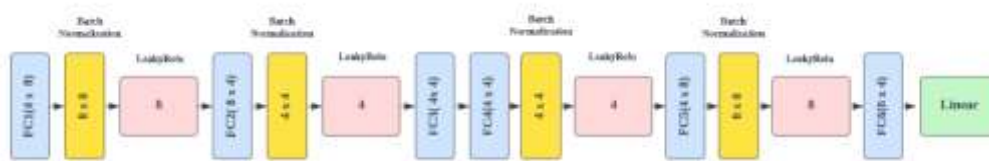
The feature extraction process plays a vital role in transforming the lengthy, noisy, and code-mixed unstructured travelogues into a structured and informative dataset. Through an intense preprocessing pipeline, each token in the travelogues undergoes annotation using our custom-created Part of Travelogue Tagger (POTT), a specialized tool designed exclusively for this research work. This tagging process categorizes each token into one of the feature categories, which includes Travel Type (TT), Travel Mode (TM), Location Climate (LC), Location Type (LT), and specific destinations (L). By annotating each token with the relevant feature category, we create a structured dataset that captures crucial information related to different aspects of travel experiences. Table 2 show the skeleton of structured dataset with essential features. This dataset becomes the foundation for training our personalized travel recommender model, enabling it to make accurate and tailored recommendations based on users' preferences and interests. The integration of the POTT tool ensures that the extracted features are aligned with the context of travelogues in the Malayalam language, enhancing the model's ability to understand the nuances and complexities of travel-related content. Overall, the feature extraction process facilitates the transformation of unstructured travelogues into a rich and organized dataset, empowering our model to deliver personalized and relevant travel recommendations for users.

**Table 2.** Structure of dataset prepared from unstructured travelogue.

Sl. No	Climate	Travel_type	Location_type	Travel_mode	Locations
1	മഴ	തനിയെ	സാഹസികം	കാർ	മണാലി
2	വെയിൽ	സോളോ	സിറ്റി	ബസ്	ദൽഹി
3	മഞ്ഞ	കുടുംബം	ഹൈറേഞ്ച്	ബൈക്ക്	ഇടുക്കി
4	തണുപ്പ്	കൂട്ടുകാർ	സിറ്റി	തീവണ്ടി	കോഴിക്കോട്
5	ചൂട്	ഫ്രണ്ട്സ്	പ്രകൃതി	വിമാനം	കശ്മീർ
6	സമ്മർ	ചങ്ങാതി	സാഹസികം	ഐക്സ്	ലഡാക്ക്
7	വിന്റർ	സഹപ്രവർത്തകർ	പ്രകൃതി	കപ്പൽ	കോവളം
7	ശൈത്യം	ഓഫീസ്	സാഹസികം	റോഡ്	ഗോവ
8	വസന്തം	ഭാര്യ	ഹൈറേഞ്ച്	കടൽ	വയനാട്
9	മഴ	ഭർത്താവ്	ചരിത്രം	ബോട്ട്	കാസർഗോഡ്
10	വെയിൽ	സഹോദരങ്ങൾ	പ്രകൃതി	സൈക്കിൾ	ആലപ്പുഴ
11	മഞ്ഞ	അമ്മ	സിറ്റി	നടന്ന്	ബാംഗ്ലൂർ
12	തണുപ്പ്	അച്ഛൻ	തീർത്ഥാടനം	സ്കൂട്ടർ	മൈസൂർ

**7. Training of autoencoder and Self-Supervised Learning Framework**

A custom autoencoder architecture has been constructed and trained using pre-processed Malayalam travelogue data, containing the decoder with an encoder. Transformation of the input features into a compact latent space representation with reduced dimensions is the duty of encoder, and reconstruction of the original data from this compressed data is the responsibility of decoder.

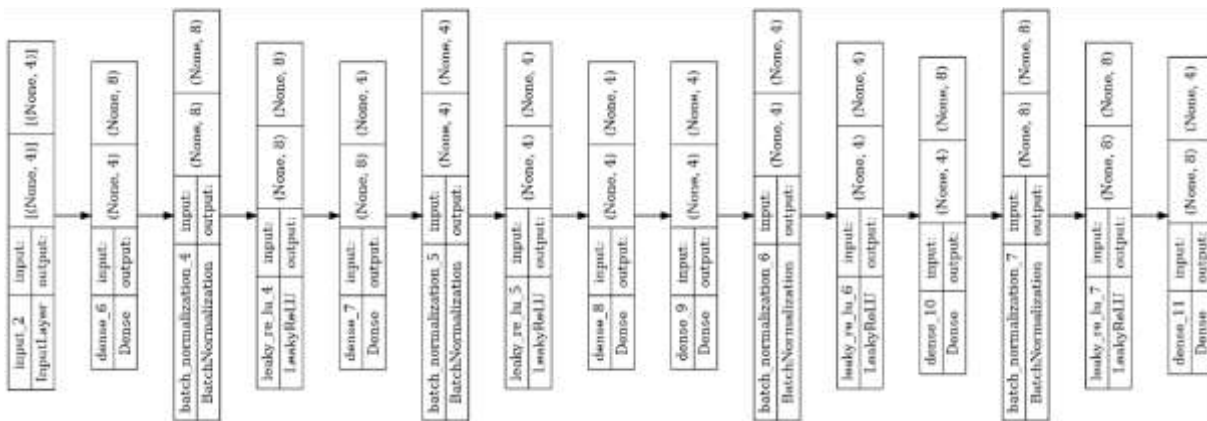


**Fig. 4** Autoencoder Model Architecture Without Compression

The autoencoder architecture shown in Figure 4 comprises an encoder network that processes the input data through two fully connected layers. The first layer, consisting of a dense layer (4x8), batch normalization (8x8), and LeakyReLU activation, reduces the dimensionality and extracts meaningful features from the input data. The output of this layer then undergoes another dense layer, batch normalization, and LeakyReLU activation in the second layer, further capturing intricate patterns in the data, resulting in generation of compressed form of input travelogue. For decoder network, it mirrors the structure of the encoder in reverse order. The first decoder layer, which includes a dense layer, batch normalization, and LeakyReLU activation, expands the dimensionality of the compressed data. The output of this layer then goes through another dense layer, batch normalization, and LeakyReLU activation in the second layer, further expanding the dimensionality and

capturing more details. The final output layer, defined as a dense layer with linear activation, aims to reconstruct the original input data.

During the subsequent phase, we formulated an autoencoder model by employing the compressed data. The encoder's structural design involved two dense layers with batch normalization and leaky ReLU activation. The initial dense layer encompassed twice the number of neurons present in the input features, while the second dense layer maintained the same neuron count, effectively achieving dimensionality reduction and capturing essential features. The decoder architecture mirrored the encoder's arrangement, comprising two dense layers with batch normalization and leaky ReLU activation, culminating in an output layer featuring an identical number of neurons as the input features and employing a linear activation function. The autoencoder underwent 50 epochs of training with a batch size of 16, with a primary focus on minimizing input-output disparities for robust data reconstruction. Furthermore, the establishment of an encoder model facilitated the extraction of the compact data representation, achieved by specifying the input and bottleneck layers.

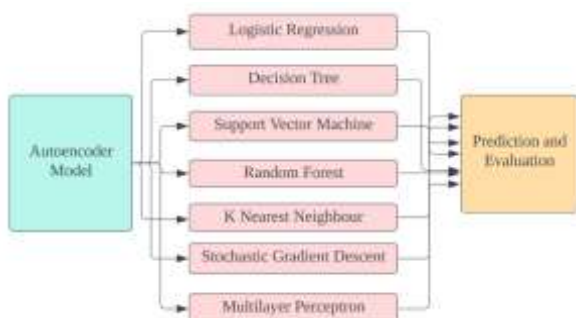


**Fig. 5** Autoencoder Model Architecture Without Compression

The model architecture depicted in Figure 5 is specifically designed for a unique task of compressing textual data in the Malayalam language, employing a pre-processed dataset rather than a synthetic one. Both the encoder and decoder components of this architecture consist of two dense layers each, integrating leaky ReLU activation and batch normalization techniques. The bottleneck layer is configured to have a size half of the input features, enabling the generation of a condensed representation of the textual data. To optimize the model's performance, it is compiled Adam as the optimizer, loss function as MSE, and accuracy. The main objective of this autoencoder is to learn a compressed representation of the textual data while capturing significant characteristics, including context, tone, semantics, and language intricacies specific to the Malayalam language. Additionally, the data is scaled using MinMaxScaler to enhance the model's efficiency and ability to handle variations in the textual input.

**9. Fusion of autoencoder with various machine learning approaches**

Next phase is to perform the efficiency of various machine learning algorithms to predict the destination as per given dataset. The output of Autoencoder model will be given as the input feature to these algorithms. Figure 6 shows the diagrammatic representation of this fusion.



**Fig. 6** Combination of Autoencoder with various ML algorithms.

## 10. Evaluation and Performance Metrics

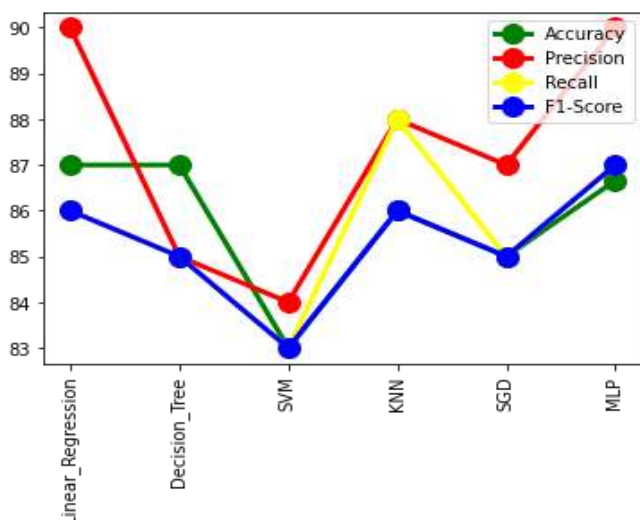
For the experiment, the considered ML algorithms are Logistic regression, Support vector machine, decision tree, Random Forest, K Nearest neighbour, SGD and MLP. Each of them have their own advantages and efficiency features. The general features are stated here. Logistic Regression is used for binary classification problems. Support Vector Machine finds the optimal hyperplane separating different classes; Decision Trees divide the dataset into subsets based on attribute values; Random Forest is an ensemble learning method utilizing multiple decision trees; K-Nearest Neighbors (KNN) classifies based on the majority of its neighboring data points; Stochastic Gradient Descent (SGD) minimizes the cost function by iteratively updating the coefficients; and a Multi-Layer Perceptron (MLP) is a type of feedforward artificial neural network consisting of multiple layers of nodes.

The experiment has done on these algorithms and the observed results are given below. While considering accuracy as the measure, all machine learning models exhibit relatively high accuracy rates, ranging from 83% to 90%. Logistic Regression and MLP achieve the highest accuracy of 90%, while SVM has the lowest accuracy of 83%. The difference in accuracy between the models is not substantial, suggesting that all models perform reasonably well in making correct predictions. Discussing about Precision, Logistic Regression and MLP models demonstrate the highest precision of 90%. The Decision Tree and SVM models have slightly lower precision scores, ranging from 84% to 85%. The precision scores show the trade-off between true positives and false positives for each model.

**Table 3.** evaluation of performance of autoencoder and ML algorithms

Sl.No.	Methodology	Accuracy			
1.	Encoder Without Compression	95.84			
2	Encoder With Compression	96.96			
Machine Learning Models with Input as Autoencoder model with compression					
Sl.No.	ML Model	Accuracy	Precision	F1 Score	Recall
1	Logistic Regression	86.96	90	86	86
2	Decision Tree	87	85	85	85
3	SVM	83	84	83	83
4	Random Forest	86.66	87	86	87
5	KNN	86	88	86	88
6	SGD	85	87	85	85
7	MLP	86.66	90	87	87

When comparing the F1 Score of these algorithms, MLP stands out with the highest F1 Score of 87%, indicating a balanced performance between precision and recall. The other models show F1 Scores ranging from 83% to 86%. The F1 Scores demonstrate the models' overall performance, considering both false positives and false negatives. Recall: KNN achieves the highest recall score of 88%, indicating its ability to correctly identify a larger proportion of positive instances. The SVM model has the lowest recall score of 83%, suggesting that it struggles to identify a significant number of positive instances. Table 3 gives the tabular details and Figure 7 gives pictorial representation.



**Fig. 7** Representation of performance evaluation of ML algorithms.

Available online at: <https://jazindia.com>



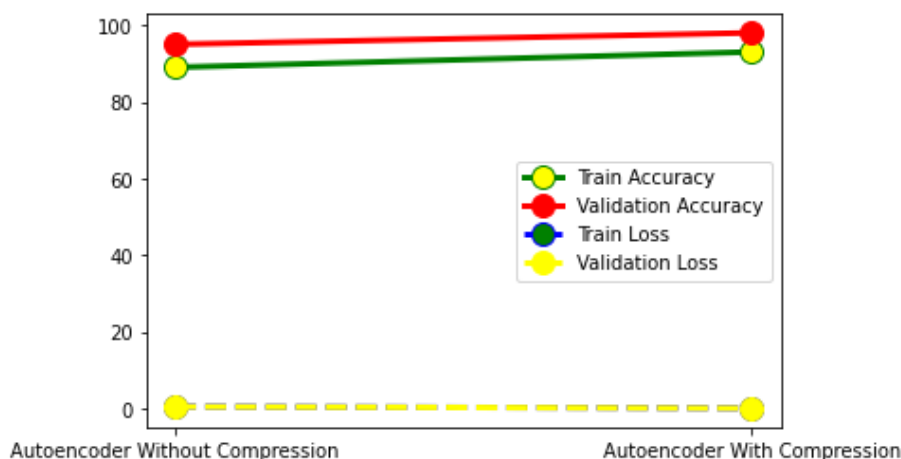
### 11. Discussion and Comparative analysis

Autoencoder is executed in two methodologies, one with compression and another without compressing data and features. Both have different results to show. Encoder Without Compression, methodology achieves an accuracy of 95.84%, which is quite high. It involves using the autoencoder to learn a compressed representation of the input data without applying further dimensionality reduction. The encoder captures essential features and patterns while preserving most of the original information.

**Table 4.** Autoencoder performance. With and without compression

Autoencoder without compression	
Loss – Training loss and validation	Training accuracy and validation accuracy
Autoencoder with compression	
Loss – Training loss and validation	Training accuracy and validation accuracy

While analysing performance of Encoder with Compression, this method achieves a higher accuracy of 96.96%. It involves using the autoencoder to further reduce the dimensionality of the input data, creating a more compact and compressed representation in the bottleneck layer. This compressed representation retains crucial information while discarding fewer essential details, resulting in a more efficient representation of the data. Table 4 shows the accuracy and loss curve of both compression algorithms as well as uncompressed algorithm. Figure 8 shows diagrammatic representation of results of models.



**Fig. 8** Performance of autoencoder model with compression and without compression.

## 12. Conclusion and Future Work

This research paper presents a novel personalized travel recommender model catering specifically to the Malayalam-speaking community. By combining an autoencoder with diverse machine learning algorithms, the model efficiently processes unstructured travelogues and extracts crucial features, including Travel Type, Travel Mode, Location Climate, Location Type, and specific destinations. The autoencoder effectively learns a compressed representation of the travel data, reducing dimensionality while retaining vital characteristics. Experimental results reveal substantial accuracy improvements compared to conventional methods. A comprehensive comparative analysis identifies Logistic Regression and MLP as the most effective models, attaining the highest accuracy and precision scores. This underscores their potential for delivering accurate and reliable travel recommendations to users. The research contributes significantly to advancing personalized travel recommendation systems, enhancing user satisfaction and engagement in travel planning. By intelligently learning from unstructured travelogues, the model provides personalized suggestions that resonate with individual preferences, fostering a deeper connection between travellers and their desired destinations.

Future endeavours focus on refining the model further by integrating user feedback and real-time data, continuously enhancing recommendation accuracy and adaptability. The study lays the foundation for more sophisticated and user-centric travel recommendation systems, ultimately enriching the travel planning experience for users within the Malayalam-speaking community.

## References.

1. M. John Basha, S. Vijayakumar, "Advancements in Natural Language Processing for Text Understanding". E3S Web Conf., 399 (2023) 04031, DOI: <https://doi.org/10.1051/e3sconf/202339904031>.
2. Mary Priya Sebastian, Santhosh Kumar G. "Malayalam Natural Language Processing: Challenges in Building a Phrase-Based Statistical Machine Translation System". ACM Trans. Asian Low-Resour. Lang. Inf. Process. 22, 4, Article 117 (2023). <https://doi.org/10.1145/3579163>.
3. R. K. Thandil, M. Basheer, and M. VK, "Natural Language Processing of Malayalam Text for predicting its Authenticity," Social Science Research Network, Jan. 2021, doi: 10.2139/ssrn.3972525.
4. D. Roy, M. Dutta, "A systematic review and research perspective on recommender systems," Journal of Big Data, vol. 9, no. 1, May 2022, doi: 10.1186/s40537-022-00592-5.
5. Jan Sawicki, et.al., "The state of the art of Natural Language Processing - a systematic automated review of NLP literature using NLP techniques". Data Intelligence 2023; doi: [https://doi.org/10.1162/dint\\_a\\_00213](https://doi.org/10.1162/dint_a_00213).
6. F. Lu and W. Zhang, "Personalized travel recommendation based on the fusion of TGI and POI algorithms". Wireless Communications and Mobile Computing, vol. 2022, pp. 1–9, Jan. 2022, doi: 10.1155/2022/4058729.
7. Guijuan ZHANG, et. al., "A survey of autoencoder-based recommender systems". Front. Comput. Sci., 2020, 14(2): 430–450 <https://doi.org/10.1007/s11704-018-8052-6>.
8. Ferreira D, et.al., "Recommendation System Using Autoencoders". Applied Sciences. 2020; 10(16):5510. <https://doi.org/10.3390/app10165510>.
9. Ding, X., et.al., "Combining Autoencoder with Adaptive Differential Privacy for Federated Collaborative Filtering". In: Wang, X., et al. Database Systems for Advanced Applications. DASFAA 2023. Lecture Notes in Computer Science, vol 13943. Springer, Cham. [https://doi.org/10.1007/978-3-031-30637-2\\_44](https://doi.org/10.1007/978-3-031-30637-2_44).
10. Guo S, Liao X, Li G, Xian K, Li Y, Liang C. "A Hybrid Recommender System Based on Autoencoder and Latent Feature Analysis". Entropy. 2023; 25(7):1062. <https://doi.org/10.3390/e25071062>.
11. Y. Bougteb, B. Ouhbi, B. Frikh, and E. M. Zemmouri, "A deep AutoEncoder-Based hybrid recommender system," International Journal of Mobile Computing and Multimedia Communications, vol. 13, no. 1, pp. 1–19, Jun. 2022, doi: 10.4018/ijmcmc.297963.
12. Kamble, Nitin, "Product Recommendation System Using Machine Learning". (2021). International Journal of Innovative Research in Computer and Communication Engineering, Volume 9, Issue 11, November 2021, doi: <http://dx.doi.org/10.2139/ssrn.4245401>.
13. Parvattikar, Suhasini and Parasar, Dr. Deepa, "Recommendation System Using Machine Learning" (2020). Proceedings of the Int. Con. on Recent Advances in Computational Techniques, doi: <http://dx.doi.org/10.2139/ssrn.3702439>.