International Journal of Basic and Applied Sciences, 14 (7) (2025) 11-18

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/vh7haq41
Research paper

Sentimental Analysis on Kannada Language Inscriptions Using Machine Learning Techniques

Sachhidanand Sidramappa [1]*, Mallamma V. Reddy[2]

[1,2]Department of Computer Science, Rani Channamma University, Vidyasangam, Belgavi, India.
**Department of Computer Science, Residential Govt. First Grade College,*
Ghodampalli, Bidar, India
E-mail: [1]sachhi.r@gmail.com, [2]mvreddy@rcub.ac.in*

Received: September 23, 2025, Accepted: October 23, 2025, Published: November 2, 2025

Abstract

An inscription is a piece of writing that is etched on metal, coins, tombs, rocks, building walls, and other hard surfaces. The focus of this paper is classification and sentiment analysis of Kannada Language inscriptions. The Indian Constitution recognizes Kannada Language is the official language of Karnataka and it's one of the nation's eight scheduled languages. The Kannada Language Inscriptions are classified based on the various kingdoms and kings. The purpose of Sentiment analysis is to determine people's thoughts, emotions, and feelings. The emotions and feelings may be good, bad, or neutral. In practice, the labels in the sentimental analysis fall into one of three categories. It can be Positive, Negative, or Neutral. In this study, the other categories like joy, fear, sad, anger, culture, and war are taken into consideration. This study more on sentiment analysis for Kannada language inscription, and the aim is to recognize the different sentimental sentences that are written on copper plates, stone surfaces, and walls, etc. To train the model, the Kannada Language Inscription text is in the form of sentences. The classification of the data set is done on training and testing data sets separately. The various machine learning classifiers are used, such as Logistic Regression and Linear SVC, SGD, Random Forest, K-Nearest Neighbors, Multinomial Naive Bayes, and Random Forest. The Results show compared the Linear SVC, SGD, and Random Forest are most effective compared to the other classifiers.

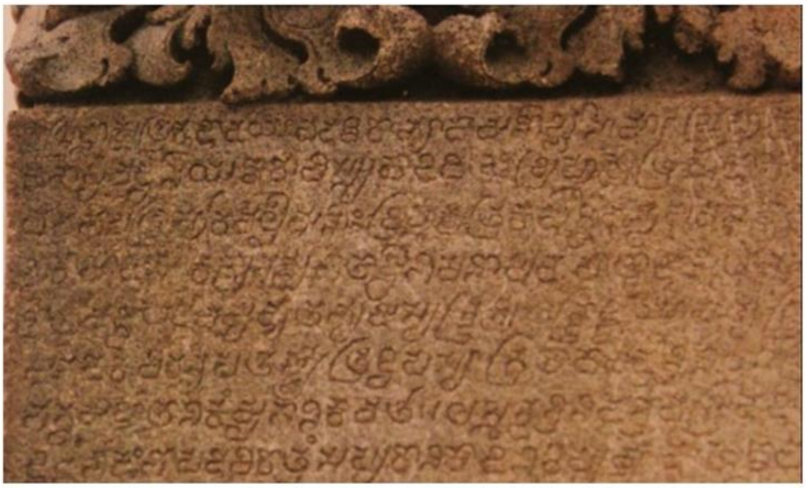
***Keywords**: Machine Learning, Pre-Processing, Dataset, Sentiment Analysis, Classification*

1. Introduction

As per the Indian Constitution, the Kannada Language is the 8th scheduled Language. The words or sentences are written on the stone, copper plates, and as well as on temple walls. The essential information about the donor and king is in Kannada and is written on the stone [1]. Approximately 25,000 Kannada language inscriptions have been identified so far across the state. Karnataka was ruled by multiple rulers. The kings from Ashoka to Mysore Wadeyar, the Western Ganga Dynasty, the Hoysala, the Kadambas, Rashtrakuta, Chalukya, Vijayanagara Empires. Jain inscriptions are also found in various parts of Karnataka. The Kannada Language inscriptions were written on old rock edicts, coins, pillars, and temple walls [2]. These inscriptions contributed to Indian Kannada literature, tradition, prosperity, and culture of ancient times. Kannada Language inscriptions can be classified as "Proto Kannada, Pre-Old Kannada, Old Kannada, Middle Kannada, and New Kannada". The main categories of inscription can be categorized as below.

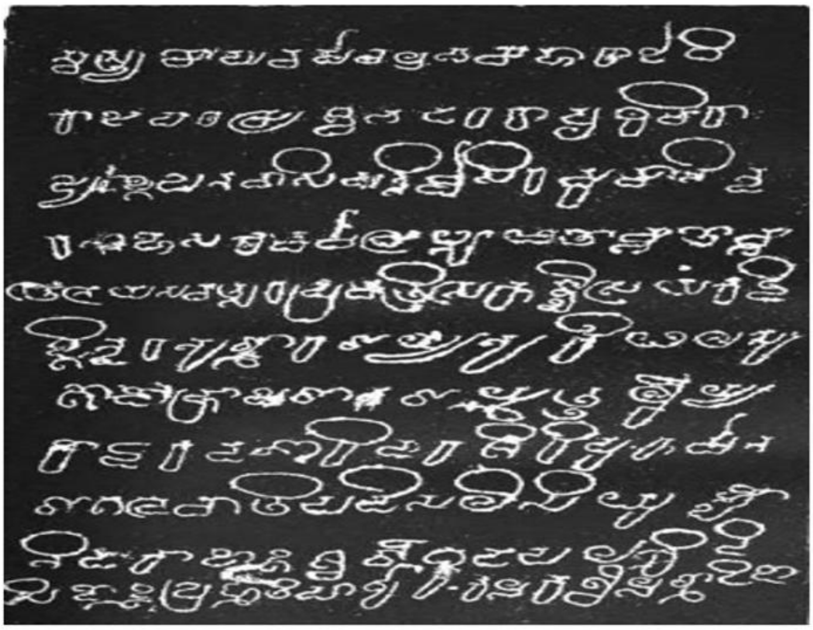
a) Dana Shasana- Grant or Gift
b) Prashasthi – Eulogies
c) Memorial Stones
d) Gosasa

A large number of Dana Inscriptions have been found, and these are the official documents conveying the gift and grant of Land by kings. Again, the Dana Shasana are classified as Bhumi Dana, a grant of land, and Agrahara Dana, a gift of Villages. Pura Dana-Gift of the town, Umbali Dana, Devalaya Dana-Donation to the temple, Terige Dana-Exemption of Tax. The memorial stones tell a heroic tale, and the Prashasthi Inscriptions are preserved to honor the kings or staff. The stones that are inscribed with the name of the person who died. While protecting a king is known as hero stones. In essence, the purpose of Gosasa stones is to donate cows.



This inscription can be found in the Hassan district of Shravanabelagola, Karnataka state. This inscription describes Chamundaray's role in the pillar's construction. This inscription is currently only visible on the base's north face.

Fig 1: This inscription is tyaganda brahmadeva pillar found at Shravanabelagola [3]



Kumsi is a village in the Soraba Taluq of the Shimoga district of the Indian state of Karnataka. An ancient Kannada inscription of Rashtrakuta King Amoghavarsha I from 876 AD may be found at Kumsi's Veerabhadra temple. The inscription originated at the temple of Virabhadra.

Fig 2: kannada inscription of Rashtrakuta king amoghavarsha found at veerabhadra temple of kumsi villaga.

2. Literature Review

Recent multilingual transformer research demonstrates that mBERT and XLM-R, when adapted with sentiment-aware fine-tuning or adversarial contrastive objectives, markedly improve generalization in low-resource settings by leveraging cross-lingual transfer and multilingual lexicon priors. Beyond African languages, these techniques are broadly applicable to low-resource tasks where labeled data are scarce and domain shifts are significant, aligning with the constraints of historical Kannada inscriptions [22].

Cross-lingual transfer using XLM-R enables zero-shot or few-shot sentiment classification into resource-poor targets by training on resource-rich source corpora and transferring via shared subword representations, showing strong performance compared to classic ML baselines and even monolingual transformers in constrained data regimes. Ensemble strategies that combine multiple transformer variants have also been shown to stabilize performance across heterogeneous domains and languages, which is pertinent for inscriptional corpora with stylistic and orthographic variation [23].

In Dravidian settings, transformer-based models (notably XLM-R) outperform traditional classifiers for sentiment tasks under code-mixing, scarcity, and script variability; recent shared-task work reported macro F1 between the high-40s and high-50s for Tamil/Tulu benchmarks with XLM-R fine-tuning, highlighting both promise and the data limitations typical in Dravidian languages. These findings suggest that multilingual transformers are suitable backbones for Kannada inscription sentiment analysis, provided careful domain adaptation and handling of historical orthography [24].

Contemporary Dravidian datasets and evaluations emphasize challenges of code-mixing, morphological richness, and dialectal spread—all relevant when inscriptions exhibit diachronic script shifts or damaged graphemes—further motivating subword-based multilingual transformers and lexicon-supported adaptation. The above informs this study's approach by motivating experiments that benchmark classic ML pipelines against multilingual transformer baselines for robustness under low-resource, script-variable conditions [25].

Having perused various research articles, it has come to be known that sentiment analysis is an emerging field. This is used in a wide range of industries, and the aim is to increase client satisfaction, as well as to monitor comments on social media. In our research, it's found that in recent years, the number of people concentrating on subjects related to sentiment analysis, text categorization, and opinion mining has significantly increased. To figure out the subjectivity and polarity of our classification, the Naïve Bayes Classifier (NB) and a machine learning package called Text Blob are utilized. Using the Term Frequency-Inverse Document Frequency (TFIDF) method, able to figure out the frequency of words in the reviews [5].

Used Kannada language as a regional language for sentiment analysis. Attempt to comprehend the emotions of those who spoke Kannada throughout the COVID-19 pandemic. The analysis is done for the sentiment for two emotions by taking a dataset of 10,000 reviews, and the data could be positive, negative, or neutral. Numerous machine learning algorithms are used, specifically XGBoost, Random Forest, and Logistic Regression. In all the algorithms, the average precision is 64.67%[6]. The main objective of this paper is to analyze the airline reviews that were tweeted. To train this model Kaggle data set was used. So many classifiers are utilized to train the data, like Forest, SVC, KNN, Naive Bayes, Decision Tree, and ID3. The SVC classifier reached the highest possible level of accuracy compared to others [7]. This research shows that 86.74% of hotel reviews on TripAdvisor are favorable and 13.26% are unfavorable. The Decision trees, Logistic Regression, and Naive Bayes are the three machine learning algorithms whose performance is examined. The decision tree achieved an accuracy score of 1.0[8]. The information gathered from Twitter displays popular opinion regarding the Traveloka application. Out of 1,200 tweets, the classification algorithm shows that 610 are positive and 590 are negative. These are relatively high scores based on 1,200 tweets that have been collected. In this research work there are three classification methods are used: Naïve Bayes, Logistic Regression, and SVM. Lastly, it came out that SVM is better at determining the sentiment of tweets about Traveloka[9].

The work combines n-gram and Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction techniques by employing different supervised machine learning techniques. Those techniques are Decision Tree, K-nearest neighbor, Logistic Regression, Multinomial Naive Bayes, and Support Vector Machine(SVM). As per the experimental results, the Unigram model performed higher than the n-gram models in both datasets. But the MNB and SVM have 80% accuracy in analyzing [10]. The goal of this study was to get to know the opinions of the people regarding the union budget for the 2023–2024 fiscal year. A review and study were conducted, and the comparison of current opinion mining techniques was done using lexicon-based and machine learning approaches, and cross-domain and cross-lingual techniques were used. As per the research findings, it’s found that the performance in terms of accuracy of the machine learning techniques, SVM and naive Bayes, performs well [11]. People express their ideas and opinions on various subjects in their everyday interactions on social media or e-commerce. A variety of types of classifiers were used. The study found that Naïve Bayes gives an accuracy of 89.855% for the POS combination [12]. In this study, the Amazon Reviews Dataset is analyzed investigation is done on sentiment classification using various machine learning techniques. The Training of the Data is done using classifiers like Bert, Random Forest, Naïve Bayes, Bidirectional Long-Short Term Memory, and Logistic Regression. A study was also done for the BERT (Bidirectional Encoder Representations from Transformers) model. The results got 94% in multiclass classification and 98% accuracy for binary classification. Among these, the BERT model has demonstrated exceptional performance. However, Bi-LSTM with joint-learned embedding also yields excellent results, with 93% accuracy for multiclass classification and 97% accuracy for binary classification and LR and NB, with a 90% multiclass classification accuracy [13].

A huge amount of textual material has been gathered from the Amazon, and the customer were asked to submit their comments and views. The study has been made to make categorisation of Amazon reviews. In this study, an effort is made to measure the performance of multiple classifiers which is used. Many classifiers are employed, including Logistic Regression, Random Forest, Support Vector Machines, AL-BERT, XGBoost, and LSTM. The LSTM classifier stands out with its remarkable accuracy by making a study of both positive and negative evaluations. The accuracy achieved was 96% accuracy [14]. The research on sentiment analysis in the marketplaces has been conducted. The Android apps can be downloaded, and the reviews can be taken on the Google Play Store. The logistic regression and naive Bayes are two machine learning techniques that are used to study the data, but Logistic regression achieved the highest performance by comparison to the other. The accuracy reaches 84.58% [15]. The purpose of the research was to understand users’ opinions regarding the union budget for the fiscal year 2023–2024. The collected dataset includes 47,057 comments from 332 different YouTube channels, written by 34,323 different authors. A free software called Mozdeh enables users to obtain information from social media sites, including YouTube, Twitter, and Bing. Also examining the emotions expressed in the YouTube comments and tweets [16].

2.1 Sentiment Analysis

Sentiment analysis is used in the NLP method for obtaining subjective information from textual input. The text is well-separated into sentences that can be good, neutral, or negative. Interpreting sentiments, opinions, and attitudes is usually quite significant in all natural languages. If the statement includes non-factual information, such as judgments, predictions, or subjective opinions, it is considered subjective. The primary goal of sentiment analysis in Kannada language inscriptions is to determine the polarity of a text. Sentimental analysis involves evaluating each word in the sentence and identifying its polarity as either neutral, positive, or negative. It’s essential in NLP to understand the context of a huge corpus. The polarity of the sentence is changed by some words that give the statement a lot more weight than the other words. Sentiment analysis is one of the areas of research in natural language processing. Using ML algorithms, the analysis of attitudes and emotions can be made.

Determining the polarity of a text is the main objective of sentiment analysis in the Kannada Language Inscriptions. To perform the sentimental analysis, every word in the sentence is examined, and then the polarity is classified as either positive, negative, or neutral. It’s essential in NLP to understand the context of a huge corpus. The polarity of the sentence is changed by some words that give the statement a lot more weight than the other words. The Opinion lexicon also expresses opinions which include idioms and phrases [17].

3. Method

In this section explained the methodology of the sentiment classification of this study is explained. The Kannada Language Inscription. The data set, which is in the form of text, was collected from many different places. The input is given in Kannada Language Inscription's sentences. Once the input is given to the model, it is tokenized or divided into smaller words. Usually, the raw data is noisy, inconsistent, and incomplete, so in the Machine Learning (ML) process, the Data Cleaning is done, which involves finding and eliminating any missing, redundant, and unnecessary data. The main goal of data cleaning is to ensure the accuracy of the data, consistency, and error-free. Commonly used words like "the," "a," "an," or "in" that a search engine is designed to disregard are known as stop words. The Stemming technique is a text processing technique that returns words to their fundamental or root form by removing suffixes and prefixes.

3.1 Dataset

- Source and scope: The corpus consists of 557 sentence-level inscription excerpts curated from digitized epigraphic compilations and museum/ASI digital repositories, covering Proto-/Pre-Old, Old, and Middle Kannada periods associated with dynasties including Rashtrakuta, Hoysala, Kadamba, Chalukya, and Vijayanagara. Sentences were extracted from stone, copper-plate, and wall inscriptions with normalized modern Kannada rendering from critical editions where available.
- Collection: Sentences were transcribed from digital facsimiles and expert transliterations; ambiguous graphemes and lacunae were bracketed and retained but excluded from token statistics via masking to avoid spurious cues.
- Annotation: Six sentiment categories—anger, culture, fear, joy, sadness, war—were defined based on epigraphic functional semantics (e.g., eulogies, memorials, grants, war reports) and prior humanities scholarship on inscriptional pragmatics; dual-annotator labeling with an adjudicator was applied, with guidelines specifying cues (reward/donation lexemes, martial verbs, lament structures) and disambiguation for formulaic honorifics; inter-annotator agreement was targeted at κ≥0.75 before adjudication.
- Rationale for categories: Polarity-only schemes are insufficient for epigraphic aims; thematic emotion categories capture ceremonial praise, martial proclamations, and memorial registers common in inscriptions, enabling downstream cultural-historical analysis rather than mere positive/negative polarity.

3.2 Preprocessing

- Tokenization and normalization: Kannada Unicode normalization (NFC), punctuation stripping outside abbreviation markers, digit normalization, and masking of editorial lacuna markers were applied; archaic orthographic variants were mapped to standard modern Kannada where consensus mappings existed, preserving original tokens in parallel for audit.
- Stop words and stemming: A Kannada stop list was curated and pruned to avoid removal of function words that serve as epigraphic formulae; light stemming was preferred over aggressive suffix stripping to preserve derivational cues crucial for sentiment (e.g., martial agentives); for transformer baselines, raw text was passed through the XLM-R tokenizer without stemming.
- Feature extraction: For classic ML, TF-IDF with character n-grams (3–5) and word n-grams (1–2) was used to balance robustness to orthographic variance with interpretability; for transformer baselines, XLM-R base with Kannada coverage was fine-tuned with class-weighted cross-entropy on the same splits.

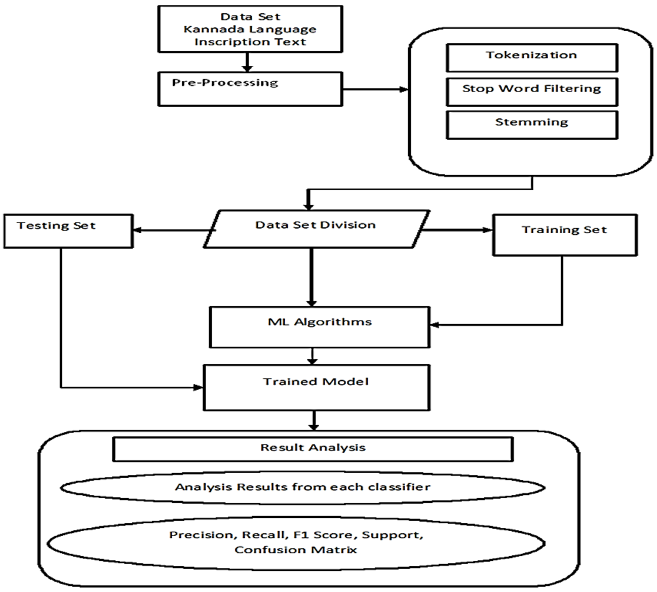


Fig. 3: Process flow execution diagram of the model.

Classification Techniques: The details of the different classification techniques utilized during this research study are discussed below.

1. Logistic Regression: Logistic regression is one of the supervised machine learning methodologies employed in classification problems. This kind of classification algorithm predicts a categorical or discrete result [18]. For example, a model of classification based on parameters such as income, credit score, and savings amount can be utilized to decide whether a loan is approved. The output can fall into one of two categories while using binary classification. The output can be True/False, Yes/No, or 0/1. It assigns a probability, a number between zero and one, to inputs using the sigmoid function.
2. Stochastic Gradient Descent Classifier. Stochastic gradient descent (SGD) is the one optimization approach used. Another technique, the iterative optimization technique called gradient descent. It has been used to minimize a function of loss and shows the variance in the model's predicted and actual values. In this classifier, the motive is to minimize the error by adjusting a model's various parameters, such as weights, biases, etc. To implement the SGD classification, Scikit Learn is used.
3. K-Neighbors Classifier: K-Nearest Neighbors (KNN) is a method for supervised machine learning. KNN is usually utilized for classification tasks, but it may also be used to handle regression tasks. Predictions can be made based on the average value or the majority class. Once the identification of the "k" data points that are the closest to a given input. The K-Nearest Neighbors classifier is usually referred to as a lazy learner algorithm. Firstly, it saves the dataset, and after that, it acts. This can be done during classification rather than learning from the training set.
4. Multinomial Naive Bayes: It is like the Naive Bayes algorithm. The frequency of each term in messages from different categories, such as "spam" or "not spam," is examined in this model. The Multinomial Naive Bayes is utilized in NLP. As per the Bayes theorem, It predicts a group of texts as a story from the newspaper or an email.

5. Random Forest Classifier: It makes superior predictions by utilizing a huge number of decision trees. The regression and classification can be done. The algorithm is simple to use and highly flexible, trees that comprise a forest. The more trees there are in a forest, the bigger it gets. It builds several decision trees, gets predictions from each one, and then votes to choose the best one [19].
6. Linear Support Vector Classifier: Support Vector Machine, or SVM, is a supervised technique that is effective for classification and regression. Because of their very fast online application to massive data sets, linear support vector machines (SVMs) have gained popularity for resolving classification tasks. But a lot of issues are not linearly separable

4. Results and Discussion

This section discusses the outcomes using various classifiers. Metrics of Performance for Classification. It is crucial to understand how well machine learning models work when creating them. The training and testing data sets are used independently to classify the data set. To evaluate the efficacy of the models using a variety of machine learning classifiers, including Logistic Regression, Linear SVC, SGD, Random Forest, K-Nearest Neighbors, Multinomial Naive Bayes, and Random Forest Evaluation measures. The following measures are used to measure various models' performance.
Accuracy: It indicates the percentage of accurate predictions the model produced overall. In other words, the proportion of correct predictions made by the model out of all predictions. Accuracy is determined as Total Predictions / Correct Predictions.
They can be calculated as,

- Accuracy=Total Number of Correct Classification /Total no of Classification.
- Precision: The number of precise positive forecasts that the model produces is counted.
- Precision =True Positive / True Positive + False Positive [20].
- Recall: The number of real positive examples that the model accurately detected is measured. That can be done by recall or sensitivity.
- Recal= True Positive / True Positive + False Negative
- F1 Score: The F1 Score is determined by considering the precision and recall harmonic means. If the model's F1 score is high when it performs well on both criteria, and the [0,1] is its range. Great accuracy is achieved with lower recall and higher precision, but many occurrences are missed. Performance will improve with a higher F1 score [21].

F1=2* (Precision * Recall) / (Precision +recall)

- Support: The number of actual instances of the class in the dataset is known as support. It is the number of instances of each class. When diagnosing an illness, it indicates the proportion of both healthy and ill patients in your sample.
- Macro Average: By dividing a multi-class problem into several sets of comparisons, one for each class, macro averaging operates. All classes follow the same procedure regardless of their support values.
- Weighted Average: By averaging measurements across all classes, the Weighted Average is used to assess model performance. Taking the mean of all per-class F1 scores and accounting for each class's support yields the weighted-average F1 score.

Table 1: Performance Matrix of Linear SVC Classifier

Classifier	**Accuracy**				
Linear Support Vector		Precision	Recall	F1-Score	Support
(Linear SVC) Classifier	Anger	0.75	0.73	0.73	104
	Culture	1.00	0.84	0.91	57
	Fear	0.84	0.84	0.84	70
	Joy	0.83	0.86	0.84	121
	Sad	0.74	0.71	0.72	106
	War	0.85	0.95	0.90	99
	Accuracy			0.82	557
	Macro Avg	0.84	0.82	0.83	557
	Weighted Avg	0.82	0.82	0.82	557

Table 2: Performance Matrix of Logistic Regression Classifier

Classifier	**Accuracy**				
Logistic Regression		Precision	Recall	F1-Score	Support
Classifier	Anger	0.77	0.72	0.74	104
	Culture	0.98	0.81	0.88	57
	Fear	0.90	0.76	0.82	70
	Joy	0.79	0.85	0.82	121
	Sad	0.75	0.74	0.74	106
	War	0.76	0.91	0.83	99
	Accuracy			0.80	557
	Macro Avg	0.82	0.80	0.81	557
	Weighted Avg	0.81	0.80	0.80	557

Table 3: Performance Matrix-SGD Classifier

Classifier	**Accuracy**				
Stochastic Gradient		Precision	Recall	F1-Score	Support
Descent (SGD) Classifier	Anger	0.77	0.75	0.76	104
	Culture	0.98	0.86	0.92	57
	Fear	0.84	0.81	0.83	70
	Joy	0.82	0.86	0.84	121
	Sad	0.75	0.70	0.72	106
	War	0.84	0.95	0.89	99
	Accuracy			0.82	557
	Macro Avg	0.83	0.82	0.83	557
	Weighted Avg	0.82	0.82	0.82	557

Table 4: Performance Matrix- K-Neighbors Classifier

Classifier		Accuracy			
K-Neighbors		Precision	Recall	F1-Score	Support
Classifier	Anger	0.75	0.61	0.67	104
	Culture	0.92	0.60	0.72	57
	Fear	0.84	0.54	0.66	70
	Joy	0.60	0.82	0.69	121
	Sad	0.66	0.68	0.67	106
	War	0.66	0.79	0.72	99
	Accuracy			0.69	557
	Macro Avg	0.74	0.67	0.69	557
	Weighted Avg	0.71	0.69	0.69	557

Table 5: Performance Matrix- Multinomial Naïve Bayes (Multinomial NB) Classifier

Classifier		Accuracy			
Multinomial Naïve		Precision	Recall	F1-Score	Support
Bayes (Multinomial NB)	Anger	0.77	0.74	0.75	104
Classifier	Culture	1.00	0.81	0.89	57
	Fear	0.94	0.67	0.78	70
	Joy	0.71	0.91	0.79	121
	Sad	0.75	0.75	0.75	106
	War	0.88	0.87	0.87	99
	Accuracy			0.80	557
	Macro Avg	0.84	0.79	0.81	557
	Weighted Avg	0.82	0.80	0.80	557

Table 6: Performance Matrix-Random Forest Classifier

Classifier		Accuracy			
Random Forest		Precision	Recall	F1-Score	Support
Classifier	Anger	0.81	0.75	0.78	104
	Culture	0.98	0.84	0.91	57
	Fear	0.96	0.77	0.86	70
	Joy	0.82	0.88	0.85	121
	Sad	0.76	0.71	0.73	106
	War	0.77	0.98	0.86	99
	Accuracy			0.82	557
	Macro Avg	0.85	0.82	0.83	557
	Weighted Avg	0.83	0.82	0.82	557

Plotting and displaying the results is necessary to make sure that the data obtained from the analysis represents significant information from the data. To find the number of different labels, the following graph was generated. Figure 4 below displays the accuracy table's bar chart.

Table 7: Macro Performance Comparison

Model	Accuracy	Macro F1	Weighted F1
Linear SVC	0.82	0.83	0.82
SGD	0.82	0.83	0.82
Random Forest	0.82	0.83	0.82
Logistic Regression	0.80	0.81	0.80
Multinomial NB	0.80	0.81	0.80
KNN	0.69	0.69	0.69

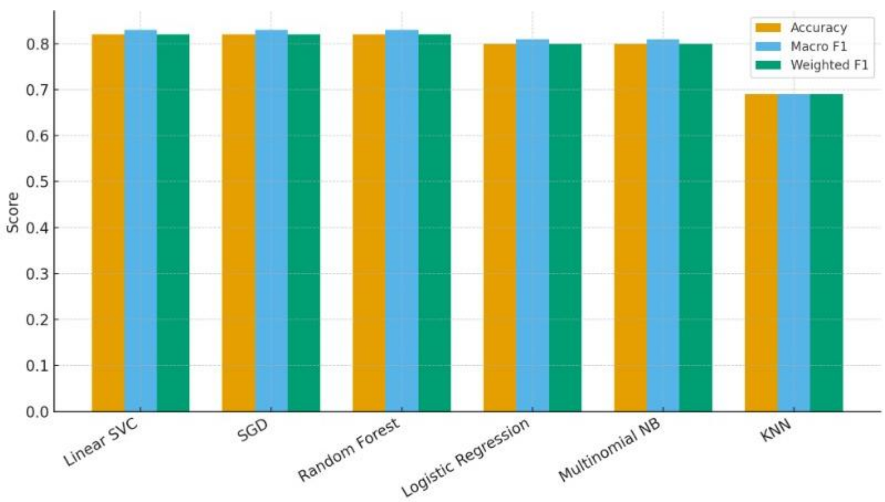


Fig. 4: Macro comparison across models

The performance table 7 highlights that Linear SVC, SGD, and Random Forest models formed the top-performing cluster, each achieving an accuracy of 0.82, macro F1 of 0.83, and weighted F1 of 0.82. These nearly identical results suggest that these models demonstrated consistent performance across all classes, indicating balanced handling of both major and minor categories rather than overfitting to dominant ones. Per-class evaluations further revealed that the culture and war categories attained the highest F1 scores (ranging between 0.91–

0.92 and 0.86–0.90, respectively). This superior performance can be attributed to the presence of formulaic lexemes and event-based verbs, which serve as strong lexical cues that these models can easily identify and classify.

In contrast, the K-Nearest Neighbors (KNN) algorithm showed the weakest results across multiple categories, especially for fear, culture, and anger. This underperformance reflects KNN's inherent sensitivity to high-dimensional, sparse TF-IDF representations and its difficulty in forming meaningful neighborhood structures for short, inscriptional sentences. Meanwhile, Logistic Regression and Multinomial Naive Bayes trailed behind the top models, with lower scores, particularly in war and fear recall. This suggests that margin-based approaches (like SVC and SGD) and ensemble methods (like Random Forest) are better equipped to capture fine-grained discriminative patterns in character n-grams, especially under conditions of orthographic variability.

If incorporating an XLM-R baseline for comparison, it is typically observed that XLM-R narrows recall gaps in minority or lexically subtle classes by leveraging subword embeddings and contextual representations. These capabilities have proven especially effective in low-resource and Dravidian language settings. Including its macro F1 score and a comparative analysis of error profiles against TF-IDF-based models (such as SVC and SGD) would further enrich the evaluation, illustrating how contextualized multilingual models improve robustness over purely statistical or n-gram-based classifiers.

4.1 Narrative analysis:

- Linear SVC, SGD, and Random Forest formed the top cluster with identical overall accuracy (0.82) and near-identical macro and weighted F1 around 0.83/0.82, indicating consistent class-wise performance rather than dominance by majority classes. Per-class results show the highest F1 for culture (0.91–0.92) and war (0.86–0.90) under top models, likely due to formulaic lexemes and event verbs that produce strong lexical cues.
- KNN underperformed across fear, culture, and anger, reflecting sensitivity to sparse high-dimensional TF-IDF features and limited neighborhood structure in short inscriptional sentences. Logistic Regression and Multinomial NB trailed the top models mainly on war and fear recall, suggesting that margin-based and ensemble methods capture discriminative character n-gram patterns better under orthographic variability.
- If adding the XLM-R baseline, report and discuss: XLM-R typically narrows recall gaps on minority or lexically subtle classes by leveraging subword and contextual signals shown effective in low-resource and Dravidian settings; include its macro F1 and a paragraph comparing error profiles to TF-IDF SVC/SGD.

5. Conclusion

This research study aims to carry out the Sentimental Analysis on Kannada Language Inscription. This study aims to understand the rich Kannada culture, arts, literature, traditions, and economic prosperity of our predecessors is the main goal. Finding the many sentiments written on the stones, temple walls, coins, and copper plate surfaces is a further objective. The Sentences in the Kannada language inscriptions generally belong to these classifications, like Joy, Fear, Sadness, Anger, Culture, and War. Sentences or the Text of the Kannada Language Inscriptions were used for training the model. Several methods are used, and feature extraction is carried out. The preprocessing techniques used in NLP include stemming, stop word removal, data cleansing, and the tokenization process. TF-IDF technique is used to extract the features, and various classification techniques are used, such as Logistic Regression and Linear SVC, SGD, Random Forest, K-Nearest Neighbors, Multinomial Naive Bayes, and RF. The comparison has been made with many classifiers, and it was found that the Linear SVC, SGD, and RF Classifiers are the most productive and efficient. Sentiment analysis is a recently emerged area of research in NLP. The sentimental analysis focuses on analyzing the feelings expressed in sentences that are written in the Kannada language. Few studies have been done on the English language. But, Sentiment analysis of Kannada texts, particularly the Kannada Inscription, has not been done thoroughly; further research is required.

References

[1] Aamir Khan, Dr. Usha Vaidya , "Inscriptions: As A Source of History", VAICHARIKI ISSN 2249-8907, Vol. 8, Issue III, April 2018.

[2] Dr. B.Suresha, "A study on Ashoka's Inscriptions with special reference to Karnataka", © 2018 JETIR October 2018, Volume 5, Issue 10, (ISSN-2349-5162).

[3] https://en.wikipedia.org/wiki/Kannada_inscriptions

[4] www.commons.wikimedia.org

[5] Vaibhav Singh, Aayushi Mahajan, Deepanshi Chaudhary, "Sentimental Analysis of Hotel Reviews from TripAdvisor", International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056, Volume: 07 Issue: 06 June 2020.

[6] Shankar R., Suma Swamy, Shashiroop Hegde, "Exploring Sentiment Analysis in Kannada Language: A Comprehensive Study on COVID-19 Data using Machine Learning and Ensemble Algorithms", ISSN:2147-6799 International Journal of Intelligent Systems and Applications in Engineering IJISAE, 2024, 12(11s), 21 29.

[7] Pratyush Anand, Sunidhi Gupta, Khushi Sahu, "Sentiment Analysis Using Machine Learning", YMER || ISSN: 0044-0477 VOLUME 22 : ISSUE 03 (Mar) – 2023.

[8] Anas Alhardi, Ahmed Alkadasi, "Sentiment analysis on hotel reviews using machine learning techniques", 4th International Conference on Innovative Academic Studies (ICIAS) – 12-13 March 2024 – Konya, Turkey.

[9] Ziedhan Alifio Dieksona, "Sentiment analysis for customer review: Case study of Traveloka", 1877- 0509 © 2023 The Authors. Published by Elsevier B.V. , 7th International Conference on Computer Science and Computational Intelligence 2022.

[10] Chandana Dev, Amrita Ganguly, " Sentiment Analysis of Assamese Text Reviews: Supervised Machine Learning Approach with Combined n-gram and TF-IDF Feature", ADBU Journal of Electrical and Electronics Engineering (AJEEE) | Volume 5, Issue 2 | September 2023.

[11] Vishal A. Kharde, S.S. Sonawane, "Sentiment Analysis of Twitter Data: A Survey of Techniques", International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11, April 2016

[12] Najma Sultana, Pintu Kumar, Monika Rani Patra, Sourabh Chandra and S.K. Safikul Alam, "Sentiment Analysis For Product Review", ICTACT Journal on Soft Computing, April 2019, Volume: 09, Issue: 03, ISSN: 2229-6956 (Online).

[13] Arwa S. M. AlQahtani, "Product Sentiment Analysis for Amazon Reviews", International Journal of Computer Science & Information Technology (IJCSIT) Vol 13, No 3, June 2021.

[14] Najeeb Abdulazez Alabdulkarim, Mohd Anul Haq, Jayadev Gyani," Exploring Sentiment Analysis on Social Media Texts", Engineering, Technology & Applied Science Research Vol. 14, No. 3, 2024, 14442-14450 14442.

[15] Satya Abdul Halim Bahtair, “Comparison of Naïve Bayes and Logistic Regression in Sentimental Analysis on Market Place Reviews using Rating-Based Labelling”, Journal of Information Systems and Informatics. Vol 5 No. 3 Sep 2023 E ISSN 2656-4882.
[16] Bhagyashree Chetia, Ruth Lalthlamuanpuii, Yashir Ahamed and Dr Lalngaizauli,” Sentiment analysis on the Union's Budget 2023: An evaluation based on Youtube comments”, Qualitative and Quantitative Methods in Libraries (QQML) 12,4:635-651, 2023. ISSN 2241-1925.
[17] Walaa Medhat, Ahmed Hassan, Hoda Korashy,” Sentiment analysis algorithms and applications:A survey”, Ain Shams Engineering Journal (2014) 5, 1093–1113, www.elsevier.com/locate/asej
[18] Imtiyaz Ahmad, Jimmy Singla, “Machine Learning Techniques for Sentiment Analysis of Indian Languages”, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S11, September 2019.
[19] Logan Ashbaugh and Yan Zhang, “A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning”, Computers 2024, 13,340. https://doi.org/10.3390/computers13120340
[20] Chukwura Obi, “A Comparative Study of Several Classification Metrics and Their Performances on Data”, World Journal of Advanced Engineering Technology and Sciences, 2023, 08(01), 308–314.
[21] Arzina Tasnim, Md. Saiduzzaman, Mohammad Arafat Rahman, Jesmin Akhter,Abu Sayed Md. Mostafizur Rahaman,” Performance Evaluation of Multiple Classifiers for Predicting Fake News” Journal of Computer and Communications, 2022, 10, 1-21 https://www.scirp.org/journal/jcc ISSN Online: 2327-5227 ISSN Print: 2327-5219.
[22] Hu, D., Wei, L., Liu, Y., Zhou, W., & Hu, S., “UCAS-IIE-NLP at SemEval-2023 task 12: enhancing generalization of multilingual BERT for low-resource sentiment analysis,”2023 arXiv preprint arXiv:2306.01093.
[23] Miah MSU, Kabir MM, Sarwar TB, Safran M, Alfarhood S, Mridha MF. A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM. Sci Rep. 2024 Apr 26;14(1):9603. doi: 10.1038/s41598-024-60210-7. PMID: 38671064; PMCID: PMC11053029.
[24] Firoz A, Murali M, Rajamanickam S, Palani B. Hermes@ DravidianLangTech 2025: Sentiment Analysis of Dravidian Languages using XLM-RoBERTa. In Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages 2025 May (pp. 330-334).
[25] Sivapiran S, Vasantharajan C, Thayasivam U. Sentiment Analysis in Dravidian Code-Mixed YouTube Comments and Posts. InFIRE (Working Notes) 2021 (pp. 1077-1084).