An AI-Driven Approach for Advancing English Learning in Educational Information Systems Using Machine Learning

Xue Peng*, Yue Wang
Tourism College, Xinxiang Vocational and Technical College, Xinxiang 453000, Henan, China

***Abstract*—In current era of globalization, English language learning is important as it has become a global language and helps people to communicate from various regions and languages. For vocational students whose main aim is to get skills and get employed, learning English for communication is important. We here present a proposed framework for learning English language which can become a foundation for a complete Artificial Intelligence (AI) based system for help and guidance to the educators. This study explores the use of diverse Natural Language Processing (NLP) techniques to predict various grammatical aspects of English language content especially focused on tense prediction which lay the foundation of English content. Textual features of Bag of words (BoW) which considers each word as a separate token and Term Frequency –Inverse Document Frequency (TF-IDF) are explored. For both diverse features, the shallow machine learning models of Support Vector Machine (SVM) and Multinomial Naïve Bayes are applied. Moreover, the ensemble models based on Bagging and Calibrated are applied. The results reveal that BoW model input for SVM and Bagging technique using TF-IDF shows optimal results with high accuracy of 90% and 89% respectively. This empirical analysis confirms that such models can be integrated with web or android based systems which can be helpful for learners of English language.**

Keywords*—*Artificial intelligence; information system; machine learning; English language learning; natural language processing

I. Introduction

The importance of the English language in today's globalized world is evident it serves as a universal medium of communication that connects people from diverse linguistic backgrounds. Fluency in English is essential for considering a broad range of information to be accessed in many ways, especially, in the institutions where many publications are in English [1]. Linguistic influence helps organize joint work and exchange results in terms of continuing the advancement of knowledge by researchers and scholars all over the world [2]. For non-native speakers, especially students in vocational colleges, the journey of learning English can be challenging, particularly when it comes to mastering grammar and sentence structure. Tenses, a core element of English grammar, pose a significant difficulty for learners, as they are essential for expressing actions in different time frames [3]. As education systems strive to provide more effective and efficient language learning, leveraging advanced technologies to address these challenges has become increasingly important. English Learning is the ability to master the capacity to write, read, speak, and comprehend the English language as well as knowledge of grammar, vocabulary, and pronunciation [4]. This learning may take place at school or university, in community classes, or through independent learning with the help of internet materials [5]. The use of English in global dimension has therefore become mandatory, and it entitles its users use in a multiplicity of sectors. In addition, cross-sectional study will ensure that the sample group is more heterogeneous and diverse regarding their grade level and learning which [6]. The importance of English learning goes beyond the need to communicate, English learning is key for success in academics and career. Knowledge in English breaks barriers and provide a drive to access information and materials since most information is found in English [7]. Many academic institutions particularly require English proficient skills to be used in reading of enhanced texts, discussions and the production of research. In addition, in today's business environments, English is often a requirement for job and promotions because of the effectiveness of the ability to facilitate cooperation with people of different teams in various countries [8]. In conclusion, learning English is not only about language mastery but is equally about improving the methods and the ways through which one can interact with the world.

The applications of AI in education, specifically in language learning, are vast and transformative. AI-powered tools can personalize learning experiences by adapting to individual learner needs, providing immediate feedback, and automating tedious tasks such as grading or content delivery [9]. In the domain of English language learning, AI can assist learners in improving their grammar, vocabulary, pronunciation, and understanding of complex linguistic concepts like tenses [10]. This personalization not only speeds up the learning process but also ensures that learners receive targeted support, thereby enhancing their chances of mastering English more efficiently [11]. Moreover, AI can bridge the gap between traditional classroom instruction and students who may not have access to high-quality educational resources, thus promoting equitable learning opportunities. The use of Education information systems in the teaching and learning of English language has emerged essential in the improvement of the teaching and learning process through use of Information and Communication Technologies (ICT) [12]. These systems provide interface to multiple learning material in form of e-books, online courses, and multimedia which enhance the reading, writing, listening and speaking ability of a learner [13] posit that the use of ICT in English Language Teaching (ELT) increases learners' attentiveness as well as encourages learner interaction with the knowledge resources and among themselves. In addition, the teachers have also described that such systems help them to give the students individual feedback and sublime courses to fit the students' needs and wants about learning [14]. By integrating these systems with the use of Recurrent Neural Networks (RNNs) and item response theory, learner's input is then subjected to the detection of context and grammatical rules used in defining the correct tenses to be used from a pool of verbs [15]. In addition to error identification, it is effective in creating practice exercises based on the learner's achievement level fostering individual learning development programs. Environmental tense predictors of language exercises that adapt automatically or provide instant feedback for right and wrong contribute to the exciting and more efficient learning modes to help learners enhance their grammatical accuracy and fluency of English language usage [16].

In this research work, we work on teaching and improving English language skills of students studying in vocational colleges as their main concern is to learn skills so that they are readily be available to the market as skilled sources. So, we have worked on basic English language skills and the use of tense which is the pivotal concept in English Grammar in English language learning by using two main AI-based technique of machine learning and state-of-the-art ensemble models including Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Bagging Classifier and CalibratedClassifierCV integrated with textual features of Bag-of-words (BoW), and Term Frequency-Inverse Document Frequency (TF-IDF), achieving highest accuracy of 90% with BoW feature when coupled with ensemble models. Further these results are evaluated using standard performance measures of accuracy, precision, recall and f1-score, showing a pathway for exploring more advanced learning abilities for vocational college students. Furthermore, contributions of this study are as follows:

- Effective Feature Selection: Used BoW and IDF features in tense prediction of tense on the English contents with the following results expressing suitability of BoW and TF-IDF for tense analysis as is relevant in classification of content.
- Comparative Model Analysis: Performed a comparison of BoW and RW performances between SVM & Multinomial Naïve Bayes; showed that SVM with BoW was 90% accurate and has outperformed other models in terms of grammar, especially tenses.
- Implementation of Ensemble Techniques: That incorporated ensemble methods including Bagging and Calibrated classifiers obtained 89% accuracy with Bagging and TF-IDF convenience features. As ensemble models enhanced the stability and the accuracy of their predictions.

The rest of the paper organization as follows: Section II presents the background knowledge of relevant literature in field of English language learning. Section III shows the comprehensive details of applied methodology including experimental setup. Section IV shares the analysis of results along with discussion. Section V provides the summary of paper in conclusion form with future directions.

II. Literature Review

The use of AI has become prominent in English Language Learning (ELL) processes as a strategy within educational information systems, improving the educational process and learners' interactions. In study [17], systematic review shows the positive and varied effects of AI technologies like ITS, NLP, and Speech Recognition on ELL. The review also notes that ITS seem to help learners the most around language proficiency or specific language skills, while speech recognition technology helps learners to improve pronunciation and speaking skills and, therefore, boosts their confidence. In addition, [18] embracing of Virtual Reality (VR) as well as Augmented Reality (AR) has changed the whole environment in which learners practice English. Another study [19] argues that these technologies foster contextually grounded environments which not only support the development of register specific language but also provide cultural context in which learners can apply the language. Appointment simulation enables learner to invariably assess their speaking performance in a supposedly realistic context, which is vital for language learning. Besides improving language skills [20], AI approaches help optimize and minimize the administrative work within the context of educational information systems. For instance, instruments like ChatGPT and education copilot help in developing courses and lesson plans, which gives such educators more time to deal with interactions and individual engaging with the student.

This automation not only decreases the possibility of grading bias but also allows instructors to give feedback promptly; instructional decisions reach a state in which it is based on the performance analytics data in real-time [21]. Still, there are obstacles in using AI in ELT because the incorporation of AI into teaching practice encounters certain difficulties. A study carried out by the Teaching English showed that schools around the world implement AI in their classrooms [22]. Similarly, concerns regarding the drawbacks of AI such as, in language use there could be bias and in terms of learning human interaction could be reduced when using the technologies [8]. Therefore, further research must reveal the effectiveness of using AI technology for learning English in the long term as well the existence of frameworks that can help educators work through its difficulties. In aggregate, the implementation of AI-inspired technologies into the scope of educational information systems is a breakthrough in learning English [11]. Introducing these systems in the classroom environments may also open the possibilities of improving the extent of teaching and learning capabilities, in the form of personalized learning experiences, coupled with the use of, for instance, immersive technologies and efficient administrative tasks. Nevertheless, the issues related to effective AI application are going to remain critical to achieve the potential benefits of utilizing AI in language instruction.

Modern progress in the areas of AI and ML have hugely impacted on improving the techniques of learning English. The implementation of some deep learning models like the CNNs and the RNNs in automating the essay grading task and appended feedbacks to the English language learners [23]. Another study, contributed to AI voice recognition in enhancing the precision on English pronunciation to help learners. Subsequent research has also examined the teaching of grammar by using AI devices to improve the learning of English grammar since the difficulty level of the material is determined by the learner's performance [24]. Authors in study [25] developed an error prediction system they say helps in teaching learners how to write by pointing out any mistake they have made and the corrective action that needs to be taken thus enhancing the learners' writing skills to carry out an enhanced learning process for writing by offering the writers feedback on grammar, vocabulary and writing style. English listening comprehension has also been a focus of machine learning, perhaps as exemplified by study [26] who posited an assessment system, incorporating the use of AI to gauge the learners' listening skills based on their proficiency level for developing applications that offer differentiated vocabulary learning according to learners' performances. Furthermore, in study [27] employed RNN for real time error correction in English as foreign language, which provided feedback for learners as soon as they wrote incorrect sentences and paragraphs so that they could correct their grammatical and writing mistakes. Lastly, NLP and machine learning to teach English syntax and semantics to learners establishing a strong foundation for learners to understand the complicated language rules [28]. These studies clearly show that English learning and teaching is the area where AI and ML are widely used as they can personalize the learning courses, develop the criterion reference assessment, and enhance the learners' skills within the different domain of the language.

III. Methodology

The following section provides a detailed procedure for specifying and categorizing understanding in the English grammar to improve teaching and learning experiences for the teachers and their learners. Concisely, the steps were carried out as Data Preprocessing, Feature Extraction methods, several Machine Learning Models, and Performance Measures have been employed as a basis for assessing the efficacy of the presented models. The, following structured approach, as shown in Fig. 1, guarantees a strong structure that responsibility helps to work out the difficulties of learning identification in English grammar.

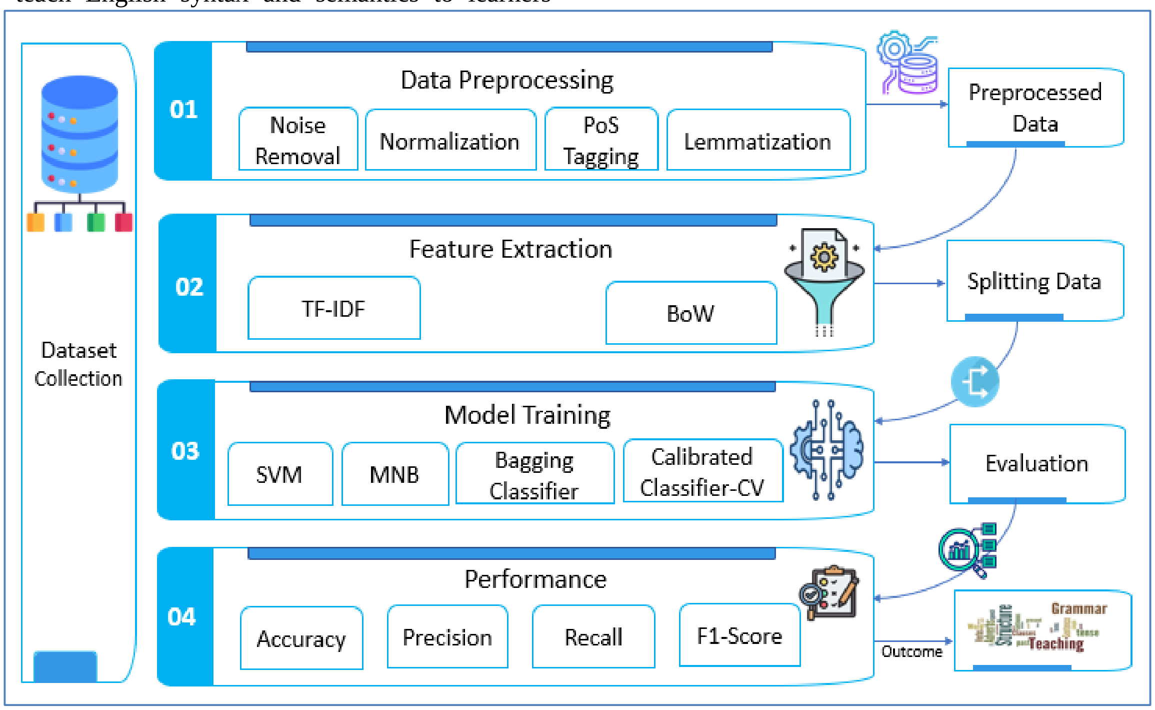

Fig. 1. The framework showing steps of the research study.

A. Data Collection and Preparation

This data set will be useful for English learners and teachers to learn and improve the usage of language in English. It includes example sentences and the tense that each of such sentences demonstrates. The data was then constructed carefully to ensure it produced sentences that clearly illustrated how the various English tenses can be used making it a rich resource for use in learning and teaching. Text preprocessing is an important step to be taken to transform textual data into NLP context. In this study, we identified several main processing procedures intended to improve the quality of the input data that is then provided to ML algorithms. There is the removal of stop-words and punctuation mark which is the first process in text mining in the process of filtering out noise words within the large datasets so that models can learn from more important words. After this, lemmatization also applied that includes the removal of prefixes and suffixes of the words and bring them to their basic form, making it easier to normalize different forms. This is particularly important for tense identification as this means that different forms of any given verb will be treated in the same way. Also, transforming text into a standard form to remove more variation from the data. However, it also employed Part-of-Speech tagging which aims at assigning a role to each word it has identified as a noun, verb, adjective and the likes. This is important especially for recognition of tenses, because verbs are of key importance in tense definition.

B. Feature Extraction

Feature engineering on the other hand is the process of extracting more meaningful features from the preprocessed text that can then be used in the actual machine learning processes. In this research, employing two primary techniques: Understanding of Term Frequency-Inverse Document Frequency (TF-IDF), and Bag of Words (BoW). Table I shows the in-depth definition of symbols used in equations.

1) Term frequency-inverse document frequency (TF-IDF): In its most simplified form, the Term Frequency (TF) can be, nevertheless, enhanced with normalization for document length differences. Normalized term frequency as is expressed by Eq. (1) can be defined as the normalized term frequency.

$$TF(t,d) = \frac{f_{t,d}}{\sum_{\bar{t}\in d} f_{\bar{t},d}}.(1+\log\left(\frac{f_{t,d}}{max_{\bar{t},d}}+1\right)) \quad (1)$$

Combining these advanced formulations, the TF-IDF score for a term t in document d relative to corpus D can be computed as in Eq. (2).

$$TFIDF(t,d,D) = TF(t,d).IDF(t,D) \quad (2)$$

Moreover, it is possible to explain the TF-IDF values based on the information gain of the terms with respect to the documents as computed using Eq. (3).

$$TFIDF(t,d,D) = \left(\frac{f_{t,d}}{\sum_{t\in d} f_{\bar{t},d}}.\left(1+\log\left(\frac{f_{t,d}}{max_{\bar{t},d}}+1\right)\right)\right).(\log\left(\frac{N+1}{n_t+1}\right)) \quad (3)$$

The mutual probability distribution can be expressed as in Eq. (4).

$$M(t;d) = \wp(t|d).\wp(d).IDF(t) \quad (4)$$

This formulation captures what we have been aiming at, in this paper, that is, capturing as to how informative each term is regarding the associated documents to be able to have an even better understanding of them within the given corpus.

2) Bag of Words (BoW): On the other hand, the Bag of Words utilized to reduce text data to a series of words while ignoring the position of words in the document and retains multiplicity. This makes word counting simple which is especially useful when trying to determine the commonly used words and particular, verb forms relating to varying tenses, by defining a weighted frequency representation that incorporates not only raw counts buts also contextual importance through various normalization techniques, defined as in equation 5. In utilizing these feature engineering strategies to make a stronger representation of the text data would be formed, which will enable the identification of the appropriate tenses in learning of the English grammar.

$$V_d = [w_{1,d}, w_{2,d}, \ldots\ldots, w_{n,d}] \quad (5)$$

Where $w_{i,d}$ is defined as in Eq. (6):

$$w_{i,d} = f_{i,d}.norm(f_{i,d}).context(t_i,d) \quad (6)$$

The procedures presented in this research for establishing an approach for constructing an automated environment for using AI to support vocational educators and learners in enhancing their mastery of English grammar and tense identification.

C. Applied Models

In the realm of NLP in dealing with issues common with English learning identification and classification. This section investigates a few more complex forms of machine learning models and how they are implemented in different ways including models like Support Vector Machines (SVM), Multinomial Naive Bayes (MNB), Bagging Classifier, and Calibrated Classifier CV. Contrasting the principles and uses of these models, to clearer understanding of how these models can be used to improve educational outcomes in English learning for both teachers and learners will be gained.

1) Support vector machine (SVM): SVM is a type of used learning method that is applied for classification problems. It does this by identifying the best hyperplane that can best separate different classes of data in a very large dimensional space. SVM works well in high-dimensional space and is not sensitive to the problem of overfitting, especially when the number of dimensions is large than the number of samples [29]. Thus, it uses a kernel size to map the data to a higher dimension so it can easily deal with non-linear relations using objective function, calculated as in Eq. (7).

$$min_{w,b,\xi} \frac{1}{2}||w||^2 + F\sum_{i=1}^{m}\xi_i \quad (7)$$

Subject of constraint to Eq. (8).

$$y_i(w.\phi(x_i)+b \geq 1-\xi_i,\ \forall i = 1,2,\ldots\ldots,m \quad (8)$$

2) Multinomial naive bayes (MNB): MNB is a probability-based classifier developed on the basic principles of Bayesian classifiers, and it is specifically useful for text classifications. It supposes that every feature is independent of other features on condition that the class is given. It is suitable for multi-class problems and most suitable with high-dimensional data such as text documents [30]. MNB performs the model by calculating the conditional probability using Eq. (9), defining each class against the words in the documents making this method simple especially for tasks like spam detection and sentiment analysis.

$$P(t|U) = \frac{P(t)P(U|t)}{P(U)} = P(t)\bigwedge_{i=1}^{n} P(v_i|t) \quad (9)$$

For each class t, the probability of observing feature vector U is given by Eq. (10).

$$P(U|t) = \frac{\bigwedge_{i=1}^{n}(f_{i,t}+1)}{\sum_{k=1}^{V}(f_{k,t}+1)} \quad (10)$$

The predicted class is determined by maximizing the posterior probability using Eq. (11).

$$\tilde{t} = argmax_t P(t)\bigwedge_{i=1}^{n} P(v_i|t) \quad (11)$$

3) Bagging classifier: This process is known as bagging – Bootstrap Aggregating which is an ensemble technique that resolves the problem of variations of a learning machine. It operates differently by building multiple models, often decision trees on different parts of the training data resulting from bootstrapping, that is random sampling with replacement based on weighted ensemble prediction, computed as in Eq. (12).

$$Y_{final} = g(\prod_{j=1}^{J} w_j h_j(x)) \tag{12}$$

The last decision is then performed by averaging or voting for these models, that Eq. (13).

$$Var(Y_{final}) = \frac{1}{J^2} \prod_{j=1}^{J} Var\left(h_j(x)\right) + (J - 1)Cov(h_j(x), h_k(x)) \tag{13}$$

4) Calibrated classifier CV: The Calibrated Classifier CV is an extension of a base classifier where cross-validation is used to enhance the probability estimation. This method refines the predicted probabilities obtained from classifiers to portray more accurate probabilities to improve decision makers in probabilistic systems, using calibrated function, defined in Eq. (14).

$$P_{calibrated}(y = 1|X) = \sigma(w^T X + b) \tag{14}$$

Where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the function used to computed objective function of calibrated model.

Cross validation is performed to make sure that calibration is done on unseen data, and thus provides higher accuracy on the prediction, computed using Eq. (15).

$$CV\left(L(w, b)\right) = \frac{1}{K} \sum_{k=1}^{K} L(w_k, b_k) \tag{15}$$

Where each fold provides a different set of parameters for calibration.

D. Performance Measure

In the context of performance evaluation of models for identification and classification for English grammar learning, several measures are used to evaluate whether models are effective in providing accurate predictions using metrics such as accuracy, precision and recall, F1 score.

Evaluation using accuracy is based on the concept of measuring the percentage probability that the model prediction for each data point is true with an overall performance, using Eq. (16).

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Prediction} \tag{16}$$

Precision, also known as positive predictive value, assesses the accuracy of the positive predictions made by the model. It is defined as in Eq. (17), the ratio of true positive predictions to the total number of positive predictions.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{17}$$

Recall, or sensitivity, evaluates to what extent the model selects the number of correct cases when it is, and in percentage of positive prediction, how accurately the model identifies the true negative cases among the wrongly predicted positives. It is defined as in Eq. (18), the accuracy of positive predictions; these are the actual number of positive predictions divided by the total number of positive predictions made.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{18}$$

F1-score is the balance between precision and recall because the harmonic mean is the better measure than average in such cases, as defined in Eq. (19). They are especially valuable in the scenario of distinguishing between the data set and the data set that is wrongly classified as the opposite class or misclassified as belonging to the opposite class by a certain model.

$$F1 - Score = \frac{2(Precision * Recall)}{Precision + Recall} \tag{19}$$

TABLE I. DESCRIPTION OF SYMBOLS USED IN EQUATIONS

Symbols	Explanation
$f_{t,d}$	Frequency of term t in document d
n_t	Number of documents containing term t
$M(t; d)$	Mutual distribution
$\wp(t\|d)$	Conditional probability
$\wp(d)$	Prior probability
V_d	Vector representation for document d
$norm(f_{i,d})$.	Normalization function to scale the frequency
$context(t_i, d)$	Contextual weighting factor shows semantic similarity measure
ξ_i	Misclassification slack variable
F	Regularization parameters
$f_{i,t}$	Frequency of feature v_i in class t
V	Vocabulary Size
$g(.)$	Majority voting function
w_j	Base classifier

IV. Results and Discussion

The descriptive analysis performed on the dataset gives an understanding of the linguistic and structural properties of the data. The Fig. 2 of distribution of tenses also represents well different tenses, and most importantly the present continuous and future tenses dominate while the tenses like future perfect continuous appear to be relatively less used. This implies a wide coverage on techniques for tension types that in turn help linguistic diversification. The analysis of the frequency of appearing of numbers of words in a sentence in Fig. 3 has indicated that most of the sentences contain between 40 and 60 words, which means that the examples used in the text should be rather appropriate for educational purposes as far as their length is concerned. The analysis of POS (Part-of-Speech) tags shows in Fig. 4 that the identified data contains a high number of nouns, verbs and determiners, and it is natural considering the focus on the sentence examples. Moreover, the word cloud of the preprocessed text graphically illustrates in Fig. 5 temporal and action-oriented words like ‘next,’ ‘year,’ ‘later’ and ‘gym’ which are evidence of time consciousness within this data set. Altogether, all these visualizations provide supporting evidence to augment the previous argument, regard to the suitability of the presented dataset for tense classification and educational purposes.

A. Results with TF-IDF

The findings from the analysis using TF-IDF accompanied by machine learning and ensemble models provide sufficient support to the proposed solution and confirm the potential of using advanced computational solutions to enhance English learning in educational systems. High accuracy of models such as SVM proved that the models can recognize and learn complex linguistic features with 88% accuracy as well as Bagging Classifier (89%) and CalibratedClassifierCV (87%) models.

These results as shown in Table II illustrating how such models can accurately identify English tenses, an essential feature of language acquisition. This way, the obtained results demonstrate the possibility of applying these models in real-life settings, including the utilization of automated grammar evaluation, individual learning environments, and language learning assistance tools. The highly favorable efficiency of the methods such as bagging Classifier proves that such an algorithm functions stably and guarantees learners receive accurate feedback regardless of the input data. Just like, SVM demonstrates a very good generalization in its modeling, which makes it suitable for distinguishing small differences in the structure of work sentences as a way of mastering English learning.

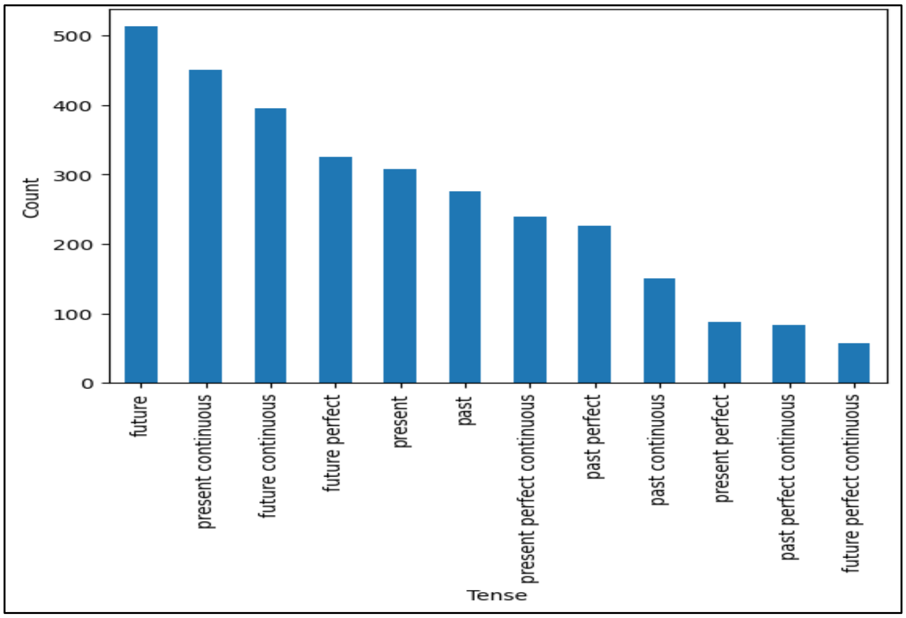

Fig. 2. Distribution of tense.

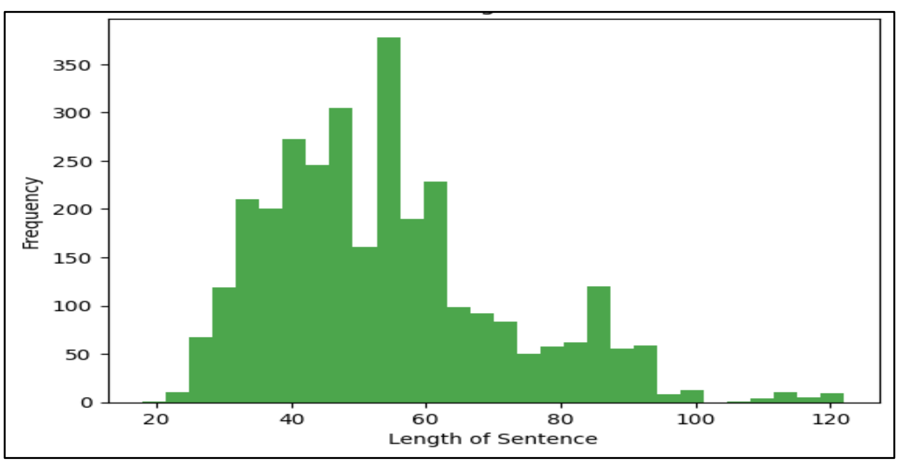

Fig. 3. Length distribution of label sentence.

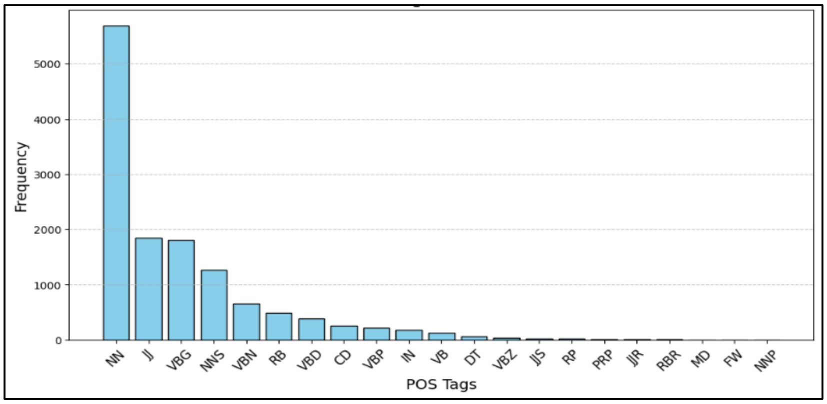

Fig. 4. POS tagging distribution.

Fig. 5. Word cloud showing most frequent words.

Furthermore, it bears mentioning such conventional and lighter weight models as MNB (80%) therefore these approaches stand for extensibility for systems of lesser computational power thereby expanding the access to language learning aids. The outcomes collectively present evidential findings of how the application of machine learning models if tuned with the TF-IDF representations can solve the problems that traditional English teaching methodologies bring in terms of its effectiveness, efficiency, and accuracy. These results support the theoretical proposition that incorporating such advanced models into educational systems can transform English language learning by making the learning process data intelligent, individualized, and precise.

TABLE II. RESULTS OF APPLIED MODELS WITH TF-IDF

Models	Accuracy	Precision	Recall	F1-Score
Shallow Machine Learning				
SVM	88	90	89	88
MNB	80	79	80	80
Ensemble Learning				
Bagging Classifier	**89**	**89**	**89**	**89**
CalibratedClassifierCV	87	86	87	85

B. *Results with BoW*

BoW when used as the feature in the classification of English tenses has highlighted the indispensability of word structures in the enhancement of English learning in educational systems. To captures the weightage categories of words, offers a more detailed picture of dependencies and structures of concrete lexemes and their contexts, which by turns allows models to grasp and disentangle tenses more effectively. The performance metrics of the models prove this approach correct, as results are briefly defined in Table III, and out of all the models created SVM attained 89% showing that it can generalize very well across syntactic structures. Bagging Classifier was the highest performing one with an accuracy of 90%. This is due to ensemble techniques where different base learners are used to improve the stability of the prediction.

This result reinforces the use of ensemble approaches in problems that involve complex discrimination and understanding of constructs, making them well suited to educational applications. On the other hand, MNB yielded an accuracy of 83% but it cannot capture complex syntactic relations. Like CalibratedClassifierCV which has 84% accuracy it is a great choice if the application that is being developed requires probabilistic outputs. Such systems can support learners by offering the right rules of learning concepts with feedback besides allowing them to take a lesson deeply into their minds through syntactic analysis. Including BoW-based models, educational platforms provide more accurate and linguistically grounded tools which contribute to progress of the learning process.

TABLE III. RESULTS OF APPLIED MODELS WITH BOW

Models	Accuracy	Precision	Recall	F1-Score
Shallow Machine Learning				
SVM	89	89	89	89
MNB	83	81	80	79
Ensemble Learning				
Bagging Classifier	**90**	**89**	**90**	**89**
CalibratedClassifierCV	84	84	84	84

The evaluation of the outcome using the TF-IDF and BoW on machine learning and related methods of ensemble brings out new perspectives of their performances, as shown in Fig. 6. In both feature extraction methods, the performances were excellent, and BoW was slightly better than TF-IDF in most of the models in terms of accuracy. Bagging Classifier achieved 90% using BoW while using TF-IDF it was 89%; for SVM BoW gives 89% while TF-IDF gives only 88%. This indicates that because of BoW's less complex representation, this model was able to capture the patterns in this dataset. Nevertheless, TF-IDF gave comparable results to the other algorithms and demonstrated good performance in evaluating term weight where the importance of terms is decisive. The two paradigm approaches emphasize on their qualities that would be useful in different areas of educational information systems for learning English.

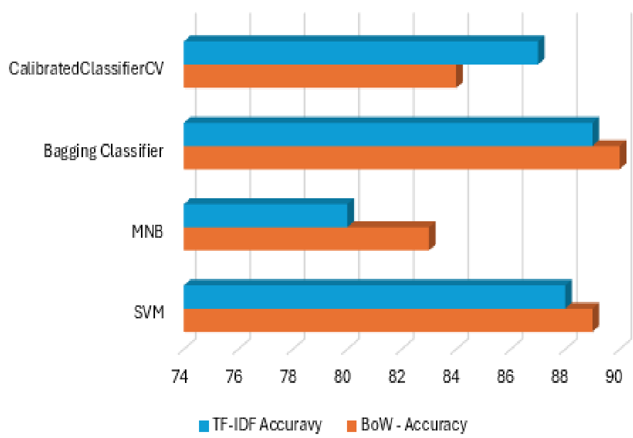

Fig. 6. Comparative analysis of applied features across accuracy measure.

V. CONCLUSION

In the modern educational system, the importance of learning cannot be overstated, as it is essential for students' academic and professional success. However, many learners, especially those in vocational colleges, face challenges in mastering key aspects of English, such as grammar, which forms the foundation of effective communication and learning abilities. Among the crucial components of English grammar, the use of tenses plays a pivotal role in ensuring clarity and accuracy in vocational college system. The advancement of AI offers significant potential to address these issues by providing personalized and scalable solutions for learning. AI-powered tools can help students understand and apply grammatical concepts more effectively, thereby enhancing their overall learning experience. Our findings in this research demonstrate the efficacy of machine learning models, particularly Support Vector Machine (SVM) and Bagging Classifiers are highly efficient for tense usage classification with accuracies of 89% and 90%, respectively using BoW and TF-IDF features. The obtained results emphasize the possibilities of development and usage of AI-technologies for improving the English language acquisition, providing a robust framework for future educational tools. Moving forward, further research can explore more sophisticated AI techniques to incorporate more complex methods for interactive learning platforms. This study will therefore create a foundation for the development of more enhanced language education to the students at vocational colleges and other institutions.

REFERENCES

[1] Li, S. (2010). Vocabulary learning beliefs, strategies and language learning outcomes: A study of Chinese learners of English in higher vocational education (Doctoral dissertation, Auckland University of Technology).

[2] Meiyan, Z. (2023). English Teaching in Higher Vocational Colleges under the Background of the. Advances in Vocational and Technical Education, 5(11), 1-9.

[3] Tividad, M. J. (2024). The English Language Needs of a Technical Vocational Institution. Mary Joy Tividad (2024). The English Language Needs of a Technical Vocational Institution. Psychology and Education: A Multidisciplinary Journal, 16(3), 256-275.

[4] Lesia Viktorivna, K., Andrii Oleksandrovych, V., Iryna Oleksandrivna, K., & Nadia Oleksandrivna, K. (2022). Artificial Intelligence in Language Learning: What Are We Afraid Of. Arab World English Journal.

[5] Chen, J. (2020, December). Strategies for improving the effectiveness of English translation teaching in higher vocational colleges based on data mining. In Journal of Physics: Conference Series (Vol. 1693, No. 1, p. 012021). IOP Publishing.

[6] Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. Computers and Education: Artificial Intelligence, 3, 100055.s

[7] Guo, L., He, Y., & Wang, S. (2024). An evaluation of English-medium instruction in higher education: influencing factors and effects. Journal of Multilingual and Multicultural Development, 45(9), 3567-3584.

[8] Soodmand Afshar, H., & Doosti, M. (2016). An investigation into factors contributing to Iranian secondary school English teachers' job satisfaction and dissatisfaction. Research Papers in Education, 31(3), 274-298.

[9] Hou, Z. (2021, February). Research on adopting artificial intelligence technology to improve effectiveness of vocational college English learning. In Journal of Physics: Conference Series (Vol. 1744, No. 4, p. 042122). IOP Publishing.

[10] Chiang, J. (2024). English Grammar Proficiency Predictability in the Prospect of Present Simple and Present Continuous Tenses: A Case from China. Teaching English Language, 18(1), 211-222.

[11] Ma, X. (2022). English Teaching in Artificial Intelligence-based Higher Vocational Education Using Machine Learning Techniques for Students' Feedback Analysis and Course Selection Recommendation. JUCS: Journal of Universal Computer Science, 28(9).

[12] Ginaya, G., Astuti, N. N. S., Mataram, I. G. A. B., & Nadra, N. M. (2020, July). English digital material development of information communication technology ICT in higher vocational education. In Journal of Physics: Conference Series (Vol. 1569, No. 2, p. 022009). IOP Publishing.

[13] Li, Y. (2020). Application of artificial intelligence in higher vocational english teaching in the information environment. In Innovative Computing: IC 2020 (pp. 1169-1173). Springer Singapore.

[14] Zinan, W., & Sai, G. T. B. (2017). STUDENTS'PERCEPTIONS OF THEIR ICT-BASED COLLEGE ENGLISH COURSE IN CHINA: A CASE STUDY. Teaching English with Technology, 17(3), 53-76.

[15] Liao, D. (2022). Deep Learning: Investigation and Analysis on the Public English Course Learning of Vocational College Students. Advances in Vocational and Technical Education, 4(2), 53-59.

[16] Jiang, H., & Wang, H. (2024, March). Designing and Implementing an Intelligent Machine Learning-Based Evaluation System for Assessing English Teaching Quality in Vocational Education. In 2024 International Conference on Interactive Intelligent Systems and Techniques (IIST) (pp. 36-40). IEEE.

[17] Manire, E., Kilag, O. K., Cordova Jr, N., Tan, S. J., Poligrates, J., & Omaña, E. (2023). Artificial Intelligence and English Language Learning: A Systematic Review. Excellencia: International Multi-disciplinary Journal of Education (2994-9521), 1(5), 485-497.

[18] Chen, Y. L., Hsu, C. C., Lin, C. Y., & Hsu, H. H. (2022). Robot-assisted language learning: Integrating artificial intelligence and virtual reality into English tour guide practice. Education Sciences, 12(7), 437.

[19] Loor, M. A. M., Solorzano, D. M. A., Katherine, A., & Moreira, V. (2024). Integration of Artificial Intelligence in English Teaching. Journal of Cleaner Producfion, 289, 125834.

[20] Amin, M. Y. M. (2023). AI and chat GPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. International Journal of Higher Education Pedagogies, 4(4), 1-15.

[21] Zhu, J., Zhu, C., & Tsai, S. B. (2021). Construction and analysis of intelligent english teaching model assisted by personalized virtual corpus by big data analysis. Mathematical Problems in Engineering, 2021.

[22] Shen, Y., Liu, Q., Zhang, K., & Zou, R. (2023, December). The Application of Artificial Intelligence Technology in Vocational College Training. In International Conference on Educational Technology and Administration (pp. 111-119). Cham: Springer Nature Switzerland.

[23] Chen, Z., Zhang, J., Jiang, X., Hu, Z., Han, X., Xu, M., ... & Vivekananda, G. N. (2020). Education 4.0 using artificial intelligence for students performance analysis. Inteligencia Artificial, 23(66), 124-137.

[24] Rane, N. L., Paramesha, M., Rane, J., & Kaya, O. (2024). Emerging trends and future research opportunities in artificial intelligence, machine learning, and deep learning. Artificial Intelligence and Industry in Society, 5, 2-96.

[25] Che, Z., Amirthasaravanan, A., Al-Razgan, M., Awwad, E. M., Mohamed, M. Y. N., & Tyagi, V. B. (2024). A Novel Renewable Power Generation Prediction Through Enhanced Artificial Orcas Assisted Ensemble Dilated Deep Learning Network. IEEE Access.

[26] Han, W. (2020). Implementing problem-based learning to enhance speaking skill in a vocational English class: an investigative study (Doctoral dissertation, Rangsit University).

[27] Wang, X., & Zhong, W. (2022). Research and implementation of English grammar check and error correction based on Deep Learning. Scientific Programming, 2022(1), 4082082.

[28] Blšták, M., & Rozinajová, V. (2022). Automatic question generation based on sentence structure analysis using machine learning approach. Natural Language Engineering, 28(4), 487-517.

[29] Shah, S. M. S., Naqvi, H. A., Khan, J. I., Ramzan, M., & Khan, H. U. (2018). Shape based Pakistan sign language categorization using statistical features and support vector machines. IEEE Access, 6, 59242-59252.

[30] Ishfaq, U., Shabbir, D., Khan, J., Khan, H. U., Naseer, S., Irshad, A., ... & Hamam, H. (2022). Empirical analysis of machine learning algorithms for multiclass prediction. Wireless Communications and Mobile Computing, 2022(1), 7451152.