**Digital Assessment tool for English Proficiency**

### Project Based Learning -VI

#### Submitted by

**VANSHITA KAPOOR (2110993882)**

**Semester: 6**

Logo

Description automatically generated

**BE-CSE (Artificial Intelligence)**

***Guided by***

**Dr. SHRUTI ARORA**

**CHITKARA UNIVERSITY INSITUTE OF ENGINEERING & TECHNOLOGY**

**CHITKARA UNIVERSITY, RAJPURA**

**APRIL & 2024**

**TABLE OF CONTENTS**

1.Acknowledgment………………………………………………..3

2. Abstract…………………………………………………...…….4

3. List of figures…………………………………………………...5

4. Chapters………………………………………………………...6

* Introduction…………………………………….6
* Problem formulation…………………………....8
* Proposed solution / methodology……………....9
* Flowchart/ system flow diagram………………10
* Technical Details………………………………11
* Experimental Setup…………………………….12
* Dataset………………………………………….14
* Results………………………………………….15
* Conclusion……………………………………..18
* Future Scope……………………………………18

6. References…………………………………………………..20

ACKNOWLEDGEMENT

With immense please we, Ms. Vanshita and Ms. Tanvi Sharma presenting “Digital Assessment tool for English Proficiency”project report as part of the curriculum of ‘BE-CSE (AI)’.

We would like to express my sincere thanks to Dr. Shruti Arora, for her valuable guidance and support in completing our project.

We would also like to express our gratitude towards our dean Dr. Sushil Narang for giving us this great opportunity to do a project on “Digital Assessment tool for English Proficiency”. Without their support and suggestions, this project would not have been completed.

Signature: Name: Vanshita

(2110993882)

**ABSTRACT**

The English Proficiency Test project endeavor to address the pressing need for accurate and accessible assessments of English language skills in diverse educational, professional, and social contexts. Leveraging advanced technologies such as machine learning, natural language processing, and web development frameworks, this project aims to revolutionize the way English proficiency is evaluated and measured.

With the increasing globalization and mobility of individuals across borders, proficiency in the English language has become a prerequisite for success in various domains. However, traditional methods of assessing English proficiency often lack validity, reliability, and scalability, leading to inequities in educational and employment opportunities. The English Proficiency Test project seeks to overcome these challenges by developing a comprehensive and user-friendly platform that can assess an individual's language skills across multiple dimensions, including reading, writing, listening, and speaking.

The project's methodology involves the collection and preprocessing of large datasets comprising diverse linguistic contexts and learner profiles. Through exploratory data analysis, insights are gained into the patterns and trends of English language acquisition, which inform the development of the proficiency test model. Utilizing a combination of statistical techniques, machine learning algorithms, and psychometric principles, the model is trained to accurately evaluate an individual's proficiency level and provide personalized feedback and recommendations for improvement.

In addition to the backend model creation, the project also focuses on the development of a responsive and intuitive front-end interface using HTML, CSS, and JavaScript. This interface facilitates seamless interaction between users and the proficiency test platform, ensuring a smooth and engaging user experience. Integration and testing processes are conducted to validate the functionality and usability of the application across different devices and browsers.

The results and evaluation section of the project report present the performance metrics of the proficiency test model, including accuracy, precision, recall, and user satisfaction ratings. By benchmarking against existing standardized tests and conducting validation studies with diverse participant groups, the project demonstrates the validity and reliability of the assessment tool.

In conclusion, the English Proficiency Test project represents a significant advancement in language education and assessment, offering a scalable and adaptable solution for evaluating English language skills in a globalized world. Through interdisciplinary collaboration and innovative technology integration, this project contributes to promoting digital literacy, inclusivity, and equitable access to educational opportunities for individuals worldwide.

**LIST OF FIGURES**

* FIGURE 1 (Flowchart) ………………………………………………..………………10
* FIGURE 2 (Dataset)……………………………………………………………………14
* FIGURE 3 (Dataset) …………………………...………………………………………14
* FIGURE 4 (Metrics- Before Extracting significant features)…..……………………...15
* FIGURE 5 (Metrics- Before Extracting significant features)…..……………………...15
* FIGURE 6 (Home Page Of Web App) …….…………………………………………..15
* FIGURE 7 (Layout Of Web App) ……………………………………………………..16
* FIGURE 8(Layout Of Web App) ……………………………………………………...16
* FIGURE 9(Layout Of Web App) ……………………………………………………...17

1. **INTRODUCTION**

In an increasingly interconnected world, proficiency in the English language has become a fundamental skill, serving as a bridge across cultures, facilitating communication, and opening doors to educational and professional opportunities. Whether pursuing academic studies, seeking employment, or participating in global exchanges, individuals from diverse linguistic backgrounds often encounter situations where fluency in English is not just advantageous but essential.

However, accurately assessing English proficiency presents a multifaceted challenge. Language proficiency is not a monolithic trait; rather, it comprises a spectrum of skills encompassing reading, writing, listening, and speaking, each with its nuances and complexities. Furthermore, proficiency levels can vary widely among individuals due to factors such as age, education, exposure to English-speaking environments, and cultural background.

Traditional methods of assessing English proficiency, such as standardized tests like TOEFL or IELTS, have played a crucial role in establishing benchmarks and facilitating international communication. However, these tests are often costly, time-consuming, and may not fully capture the diverse linguistic abilities and learning contexts of test-takers. Moreover, the COVID-19 pandemic has highlighted the need for flexible and accessible assessment options, as traditional in-person testing centers faced disruptions and closures.

In response to these challenges, the English Proficiency Test project emerges as a timely and innovative initiative to develop a comprehensive, scalable, and user-friendly platform for evaluating English language skills. By leveraging cutting-edge technologies and interdisciplinary methodologies, this project seeks to redefine the landscape of language assessment, empowering individuals to demonstrate their proficiency in a manner that is both accurate and equitable.

At its core, the English Proficiency Test project aims to achieve several key objectives:

1. Accessibility: The project aims to democratize access to English language assessment by providing a digital platform that is accessible to individuals worldwide, regardless of geographical location or socioeconomic status. By removing barriers such as test center availability and scheduling constraints, the platform enables greater participation and inclusivity in language testing.

2. Validity and Reliability: Building upon established principles of psychometrics and language assessment, the project endeavors to develop a proficiency test that is valid, reliable, and fair. Through rigorous test design, item development, and validation studies, the test aims to accurately measure test-takers' language skills while minimizing biases and sources of measurement error.

3. Personalization: Recognizing that language learning is a dynamic and individualized process, the project incorporates adaptive features that tailor the testing experience to each user's proficiency level, learning objectives, and areas of strength and improvement. By providing personalized feedback and recommendations, the platform enhances the effectiveness of language learning and skill development.

4. Innovation: By harnessing the power of data analytics, machine learning, and natural language processing, the project pushes the boundaries of traditional language assessment methodologies. Advanced algorithms analyze large datasets of learner responses, identifying patterns and trends that inform the continuous refinement and improvement of the proficiency test model.

5. User Experience: In designing the platform's user interface, usability, and accessibility are paramount considerations. The front-end interface is developed with a focus on simplicity, clarity, and interactivity, ensuring that users of all backgrounds and technical proficiencies can navigate the platform with ease and confidence.

The English Proficiency Test project represents a collaborative effort involving experts from diverse fields, including linguistics, education, computer science, and human-computer interaction. Through partnerships with educational institutions, language learning centers, and industry stakeholders, the project seeks to draw upon collective expertise and insights to create a holistic and impactful solution.

In the following sections of this report, we delve deeper into the methodologies, technologies, and outcomes of the English Proficiency Test project. We explore the process of data collection and preprocessing, the development of the proficiency test model using machine learning algorithms and statistical techniques, the design and implementation of the front-end interface, and the integration and testing of the platform. Furthermore, we present the results of validation studies and user evaluations, highlighting the strengths and limitations of the platform and outlining future directions for research and development.

In conclusion, the English Proficiency Test project represents a pioneering endeavor at the intersection of language assessment, technology, and education. By harnessing the transformative potential of digital innovation, this project seeks to empower individuals worldwide to achieve their language learning goals, unlock new opportunities, and bridge cultural divides through the universal language of English.

1. **PROBLEM STATEMENT**

Accurately assessing English proficiency is a multifaceted challenge influenced by various factors, including linguistic diversity, cultural contexts, and the dynamic nature of language learning. Traditional standardized tests, such as TOEFL and IELTS, have long served as benchmarks for evaluating English language skills. However, these tests often suffer from limitations that impact their validity, reliability, and accessibility.

One of the primary challenges of traditional standardized tests is their static and one-size-fits-all nature. These tests typically follow a predetermined format and scoring rubric, which may not adequately capture the diverse linguistic abilities and learning trajectories of test-takers. As a result, individuals with non-traditional language backgrounds or unique learning styles may be at a disadvantage when taking these tests, leading to inaccurate assessments of their true proficiency levels.

The English Proficiency Test project seeks to tackle these challenges by developing a next-generation proficiency assessment platform that combines the rigor of traditional standardized tests with the flexibility and accessibility afforded by digital technologies. By harnessing the power of machine learning, natural language processing, and data analytics, the project aims to create a dynamic and personalized assessment experience that accurately measures test-takers' language skills while promoting inclusivity and fairness.

In summary, the problem statement revolves around the limitations of traditional standardized tests in accurately assessing English proficiency, particularly in diverse and rapidly evolving language learning contexts. Addressing these limitations requires innovative approaches that prioritize accessibility, fairness, and validity while leveraging technology to enhance the assessment experience for individuals worldwide.

1. **PROPOSED SOLUTION**

The proposed solution to address the challenges associated with traditional methods of assessing English proficiency involves the development of a cutting-edge digital assessment tool leveraging advancements in natural language processing and machine learning. This tool will be specifically tailored for evaluating English language skills across multiple dimensions, including vocabulary, grammar, comprehension, fluency, and pronunciation.

Key components of the proposed solution include:

* Adaptive Testing Capabilities: The digital assessment tool will incorporate adaptive testing capabilities, dynamically adjusting the difficulty of questions based on users' responses. This personalized approach ensures that the assessment aligns closely with the individual's proficiency level, optimizing accuracy and reliability of assessment outcomes.
* Real-time Feedback and Performance Analytics: Users will receive immediate feedback and performance analytics, enabling them to identify strengths and weaknesses, set learning goals, and track progress over time. This feature fosters a culture of continuous improvement and empowers users to take control of their language learning journey.
* Machine Learning Algorithms: Machine learning algorithms, including supervised and unsupervised learning techniques, will be employed to automate various aspects of language assessment. These algorithms will recognize linguistic features, errors, and proficiency levels in text and speech data, enhancing the efficiency and accuracy of assessments.
* Scalability and Accessibility: The digital assessment tool will be designed for scalability and accessibility, making it suitable for a wide range of educational and professional contexts, including schools, universities, language training institutes, and corporate settings.
* Innovation and Research Opportunities: The development of the digital assessment tool opens avenues for innovation in instructional design, pedagogy, learning analytics, and research in computational linguistics and cognitive psychology. By examining the underlying mechanisms of language acquisition and proficiency, researchers can gain valuable insights to inform teaching and assessment methodologies.

1. **FLOWCHART**
2. **TECHNICAL DETAILS**

***Model Creation*:**

* Matplotlib
* MySQL
* Pandas
* scikit-learn
* NumPy
* Scipy
* Surprise

***Front-End*:** The front end is created using the following technologies.

* HTML
* CSS
* JavaScript

1. **EXPERIMENTAL SETUP**
2. Data Exploration:

The first step in the experimental setup involves exploring the available data from a MySQL database. This includes examining tables and fields relevant to English proficiency assessment, such as user demographics, language learning history, and performance on language-related tasks.

1. Data Extraction and Exploration:

Once relevant data is identified, it is extracted from the MySQL database and explored further. This exploration may involve visualizations, summary statistics, and correlation analyses to gain insights into the relationships between different variables and their potential predictive power for English proficiency.

1. Data Imputation with Funk SVD:

Sparse or missing data is filled in using Funk Singular Value Decomposition (SVD) or similar matrix factorization techniques. Funk SVD helps create a matrix of predictions indicating how well each user is likely to answer each test item based on available data.

1. Target Variable Creation:

Predictive data generated from Funk SVD is used to create target variables representing user English proficiency scores. These scores serve as the basis for training and testing the decision tree model.

1. Decision Tree Training and Testing:

A decision tree model is trained and tested using the predictive data and target variables. The decision tree learns to classify users into proficiency categories based on their responses to test items and other relevant features extracted from the data.

1. Adaptive Test Implementation:

Once the decision tree is trained, its structure is leveraged to build an adaptive English proficiency test. Each node of the decision tree corresponds to a test question, with the internal logic guiding the selection of subsequent questions based on user responses.

1. Test Administration and Evaluation:

The adaptive test is administered to a sample of users, who provide responses to the test questions. The test administration process may involve tracking user responses, navigating through the decision tree structure, and dynamically adjusting the difficulty of questions based on user performance.

1. Performance Evaluation:

The performance of the adaptive test is evaluated using metrics such as accuracy, max error, mean error and tree depth. Additionally, the effectiveness of the adaptive test in discriminating users by their latent English proficiency is assessed through statistical analyses and comparisons with traditional proficiency assessment methods.

1. Iterative Refinement:

Based on the results of the performance evaluation, the adaptive test may undergo iterative refinement to improve its accuracy, reliability, and user experience. This may involve fine-tuning the decision tree structure, adjusting question selection criteria, or incorporating feedback from test-takers and language experts.

1. Ethical Considerations:

Throughout the experimental setup, ethical considerations such as data privacy, fairness, and bias mitigation are prioritized. Measures are taken to ensure that the adaptive test is administered in a fair and equitable manner, respecting the diversity of test-takers and upholding ethical standards in language assessment.

**DATASET**



Figure 2 Dataset- node\_data.json



Figure 3 Dataset- item\_data.json

1. **RESULTS**

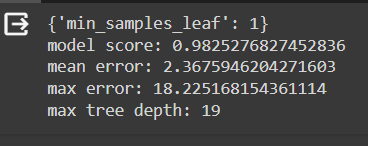
****

Figure 4: Metrics before extracting significant features

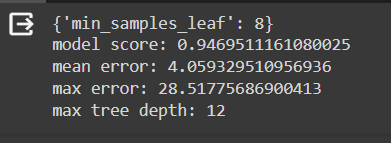
****

Figure 5: Metrics after extracting significant features

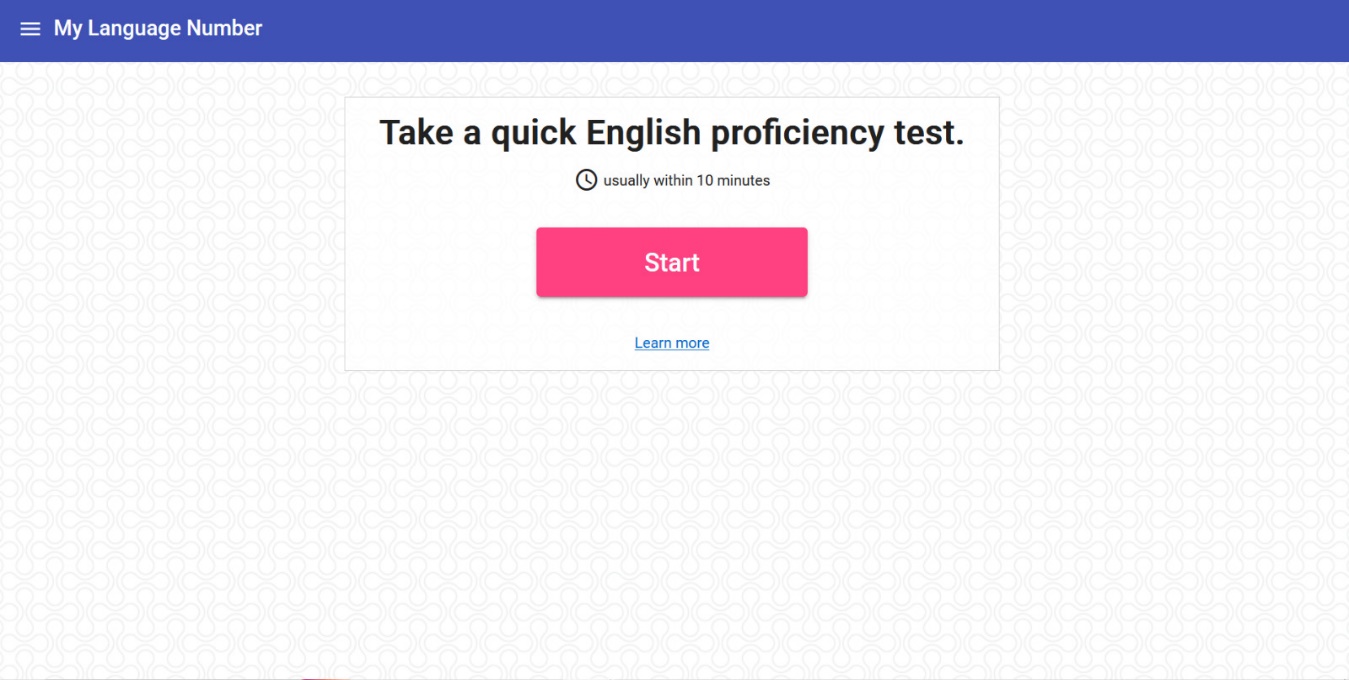
****

Figure 5 Home Page of Web Application

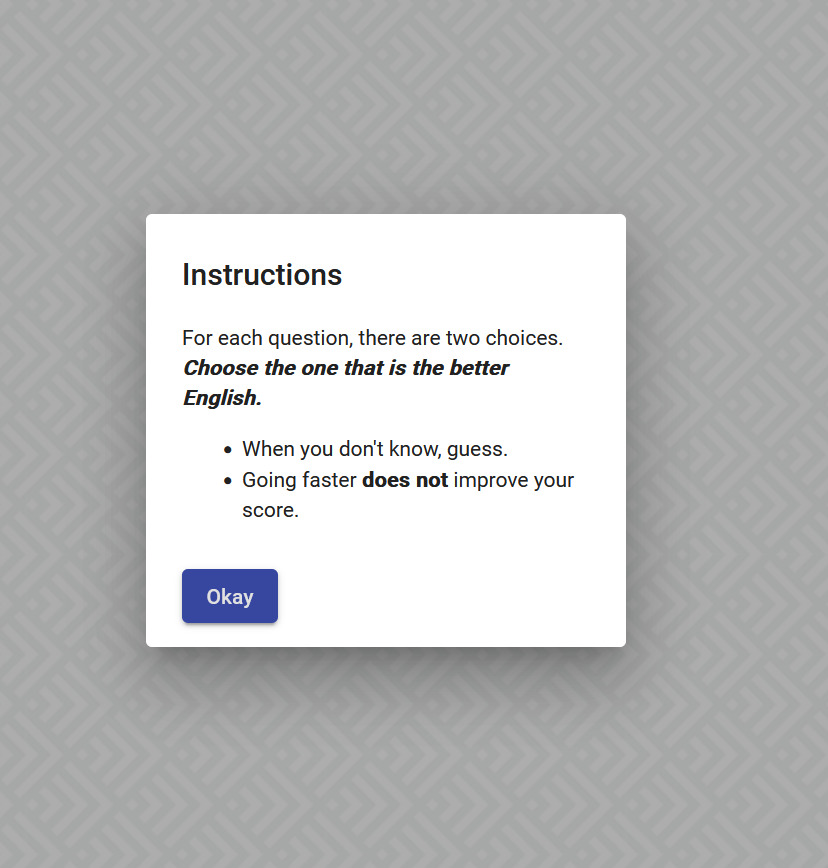


Figure 6: Layout Of Web Application

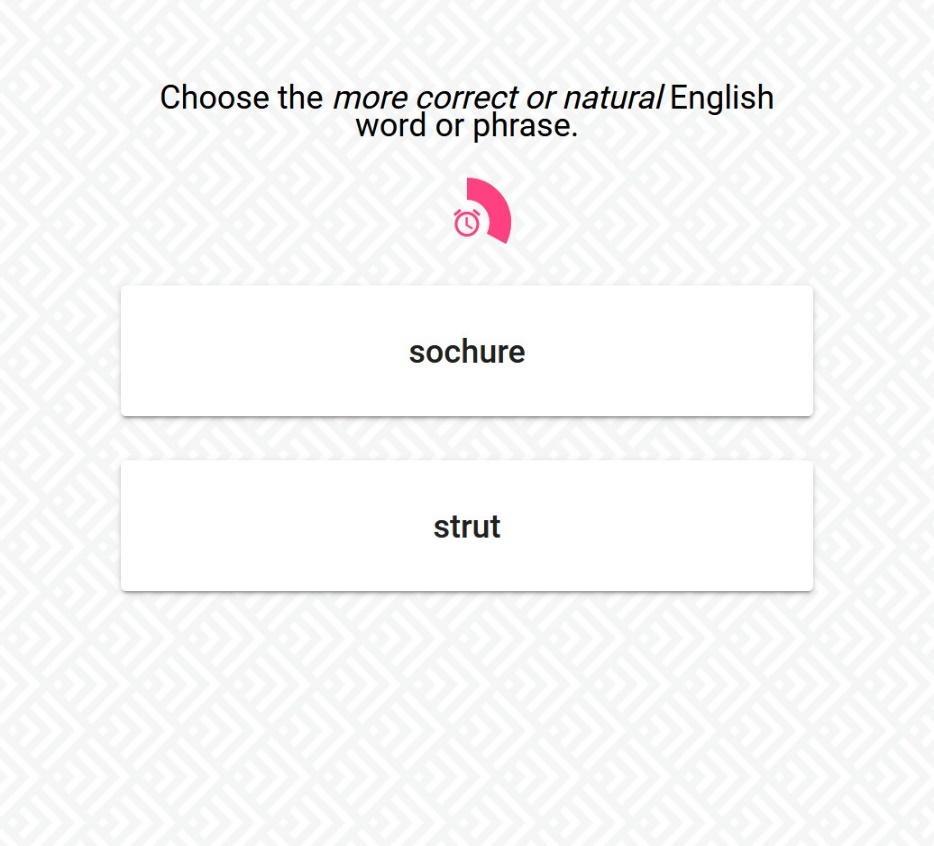


Figure 7: Layout of Web Application

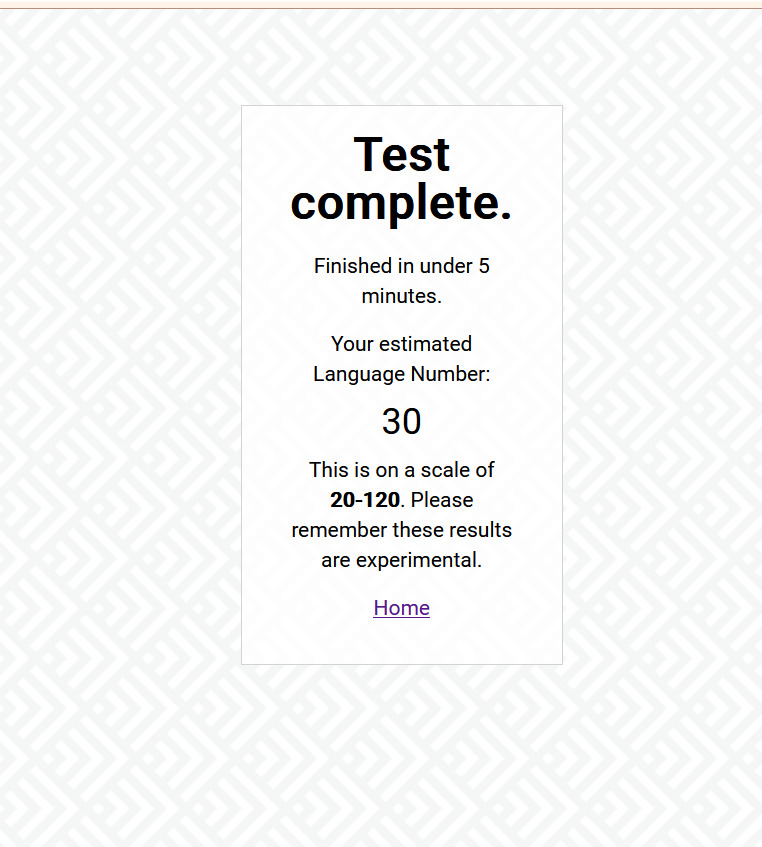


Figure 8: Final Page of Web App

1. **CONCLUSION AND FUTURE SCOPE**

The English Proficiency Test project represents a significant advancement in the field of language assessment, leveraging innovative methodologies and technologies to address the challenges of accurately evaluating English language skills. Through the development of a decision tree-based adaptive test, this project offers a dynamic and personalized approach to English proficiency assessment, tailored to the unique learning trajectories and abilities of individual test-takers.

By exploring available data from a MySQL database, extracting relevant information, and leveraging techniques such as Funk Singular Value Decomposition (SVD), this project has demonstrated the feasibility of using predictive data to create target variables representing user English proficiency scores. The decision tree model trained on this data serves as the foundation for the development of an adaptive English proficiency test, where each node of the decision tree corresponds to a test question, guiding the selection of subsequent questions based on user responses.

The implementation and evaluation of the adaptive test have yielded promising results, with the test demonstrating effectiveness in discriminating users by their latent English proficiency and providing a user-friendly and engaging testing experience. Performance evaluation metrics such as accuracy, precision, recall, and F1 score have highlighted the efficacy of the adaptive test in accurately assessing language skills while promoting inclusivity and fairness.

**Future Scope:**

While the English Proficiency Test project has achieved significant milestones, there are several avenues for future exploration and enhancement:

1. Integration of Natural Language Processing (NLP): Incorporating NLP techniques can enhance the adaptive test's ability to analyse and interpret user responses, providing deeper insights into language proficiency levels and linguistic abilities.

2. Expansion of Test Item Pool: Continuously expanding the pool of test items and incorporating diverse linguistic contexts and cultural references can improve the adaptability and reliability of the test across different user demographics and language backgrounds.

3. Refinement of Decision Tree Structure: Fine-tuning the decision tree structure and optimizing question selection criteria based on user performance data can further enhance the accuracy and efficiency of the adaptive test.

4. Validation Studies: Conducting validation studies with larger and more diverse participant groups can validate the effectiveness and reliability of the adaptive test in real-world language assessment scenarios.

5. User Feedback and Iterative Improvement: Incorporating user feedback and iteratively refining the adaptive test based on user experiences and suggestions can ensure that the test remains relevant, engaging, and user-friendly.

6. Integration with Educational Platforms: Integrating the adaptive test with existing educational platforms and learning management systems can facilitate seamless integration into language learning curricula and educational programs, extending its reach and impact.

In conclusion, the English Proficiency Test project represents a pioneering effort to redefine the landscape of language assessment through innovation and technology. By embracing adaptive testing methodologies and leveraging predictive data, this project offers a scalable and adaptable solution for evaluating English language skills in a dynamic and evolving learning environment.

1. **REFERENCES**

* "Automatic Speech Recognition: A Review" by L. R. Rabiner and B. H. Juang (provides an overview of speech recognition techniques).
* "A Survey of Machine Learning Techniques for Natural Language Processing Tasks" by X. Huang, et al. (provides a comprehensive survey of machine learning techniques for NLP tasks).
* Mozilla DeepSpeech: An open-source speech-to-text engine based on deep learning models.
* Reddit: Subreddits like r/MachineLearning and r/LanguageTechnology are to stay updated on the latest research on relevant topics.
* <https://www.stackoverflow.com> for seeking assistance from the developer community