



PARSHWANATH CHARITABLE TRUST'S
A.P. SHAH INSTITUTE OF TECHNOLOGY
Department of Computer Science and Engineering
Data Science



English Vocabulary Enhancer

Gauri Iyer - 22107030

Siddhesh N Patil - 22107019

Abhimanyu Deshmukh - 22107036

Ganesh Patil - 22107045

Project Guide
Prof. Rajashri Chaudhari

Outline

- Introduction
- Literature Survey of the existing systems
- Limitations of the existing systems
- Problem statement
- System Design
- Technologies and methodologies
- Implementation
- Conclusion
- References

Introduction

- A strong command of the English language is essential for both academic success and career advancement.
- Preparing for exams like TOEFL, GRE, or IELTS, or simply aiming to improve communication skills, vocabulary plays a critical role.
- The challenge lies in finding a systematic, personalized method for learning and revising vocabulary.

1.1 MOTIVATION :

- The need for this project arose from personal observations and discussions with peers preparing for competitive exams.
- Many individuals face difficulties in expanding their vocabulary due to lack of structure, inconsistent practice, and limited resources for effective revision.

Introduction

1.2 OBJECTIVES :

- To help students/working professionals learn new words using dynamic difficulty level shifting based on the Leitner System.
- To help them stand out in their careers with linear and step-by-step learning which is delivered by 'Linear Regression'.
- To improve one's learning abilities by increasing understanding of complex words. (Word classification will be done on the basis of difficulty using sklearn)
- To let user track his/her progress by giving stats based on their learning and practicing.

Literature Survey of the existing system

Sr No	Title	Author	Year	Outcomes	Methodology	Result
1	English Word Difficulty Classifier Based on Random Forest Model	Miao Peng , Yujie Wu , Yannan Qiu	2023	The results show that the model achieves 97.2% accuracy on the training set and 97.8% on the test set, demonstrating strong generalization. This high accuracy indicates the model's ability to predict word difficulty effectively based on the selected features.	The classifier was developed with the help of Random Forest Classifier and K-means Classifier.	Its high accuracy suggests potential applications for evaluating text readability, with opportunities for future research to apply this approach to other languages or improve predictive capabilities.

Literature Survey of the existing system

Sr No.	Title	Author	Year	Outcomes	Methodology	Result
2	Word difficulty classification prediction based on ARIMA and DBSCAN	Xilun Li , Lingyu Sun , Jigang Li	2023	Using clustering algorithms, we generated two key indicators: "sum" (representing the percentage of players guessing the word in one or two attempts) and "rank" (representing the percentage of players participating in hard mode).	Pattern finding using ARIMA, DBSCAN for identifying clusters	Future iterations of Wordle maintain a balance between difficult and easy words. Revealing some of the underlying logic or patterns in word selection could make it easier for players to anticipate and strategize, enhancing engagement without sacrificing challenge.

Literature Survey of the existing system

Sr No.	Title	Author	Year	Outcomes	Methodology	Result
3	Word Difficulty Level Prediction System Using Deep Learning Approach	Shivam Parihara, Shubhi Miradwal , Abhinav Panse , Ruchi Patel	2019	The model was trained on 1,400 words, and in cases of error (562 instances), the bidirectional neural network sent data backward through the layers to improve accuracy. The training and testing dataset division showed that the highest accuracy of 97.5% was achieved when 16.7% of the dataset was used for testing.	This system is proposed on the concept of LSTM and phonetics.(Long Short-Term Memory is a type of recurrent neural network (RNN) architecture specifically designed to learn and model sequences of data)	The model's performance was compared using LSTM and GRU cells, with LSTM proving superior. The LSTM-based model achieved an accuracy of 97.5%.

Creating the Dataset

Used BeautifulSoup (BS4) and Selenium to scrape the data from the web:

- ▶ We searched for multiple vocabulary websites
- ▶ Chose the website with good quality words and meanings
- ▶ Automated a chrome based webdriver to crawl through the website and reach at the target location using selenium
- ▶ Used beautifulsoup to scrape the words, meanings and their frequency (per how many words that word is used in normal vocabulary and conversations)

Supervised learning

Training the model

- We need to have features (weights) to train a model
- We performed feature engineering for the word-difficulty training
- As the Syllable count, Word frequency and word length all are non-categorical data, they can be used for training directly

Linear Regression

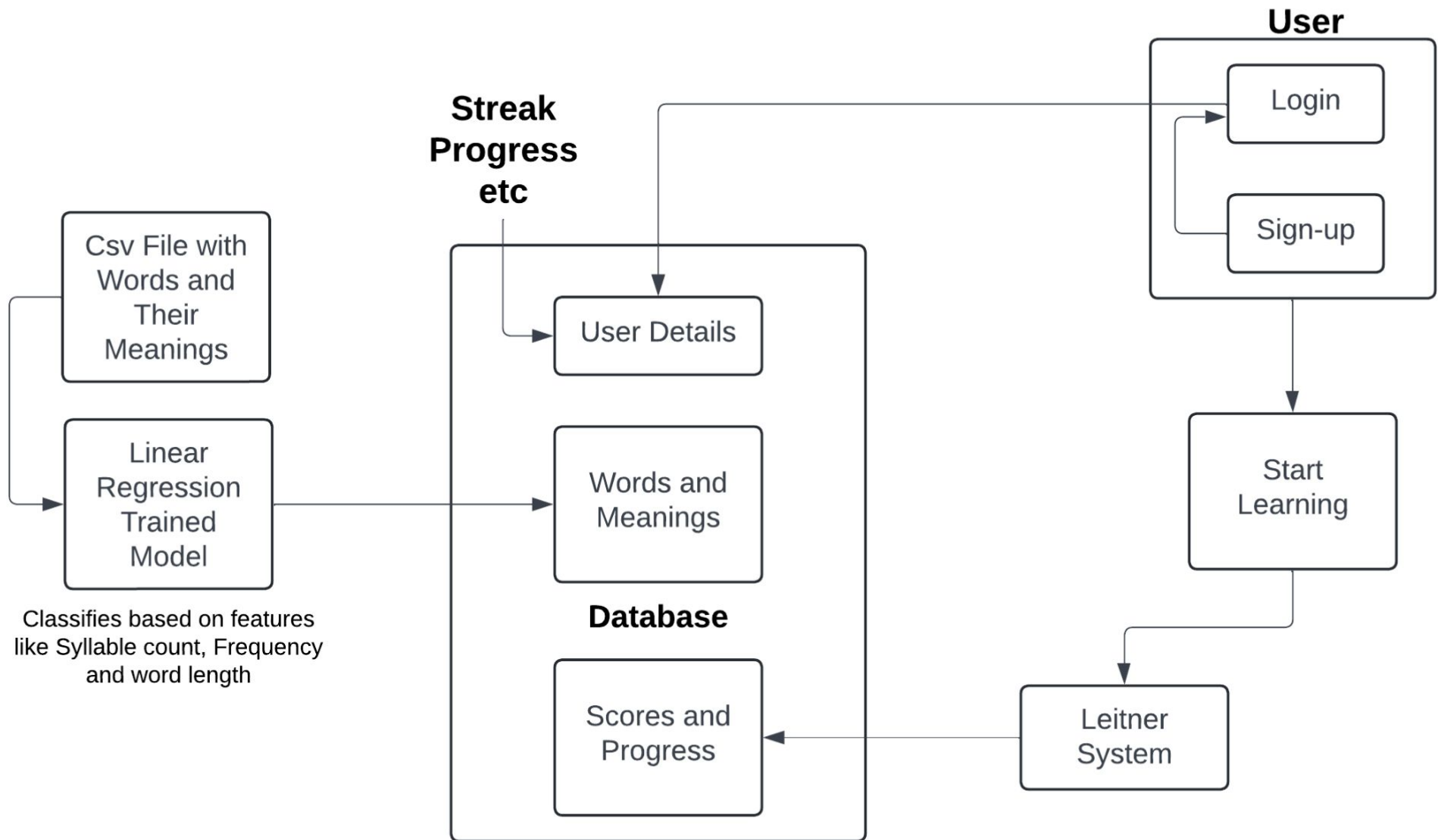
- We have 3 independent variables and 1 dependent variable and so multilinear regression becomes a good option
- Random forest classifier but it gave around 72 percent accuracy, but regression gave 91 percent accuracy

Syllable Count

Word
Frequency

Word Length

Data Flow Diagram



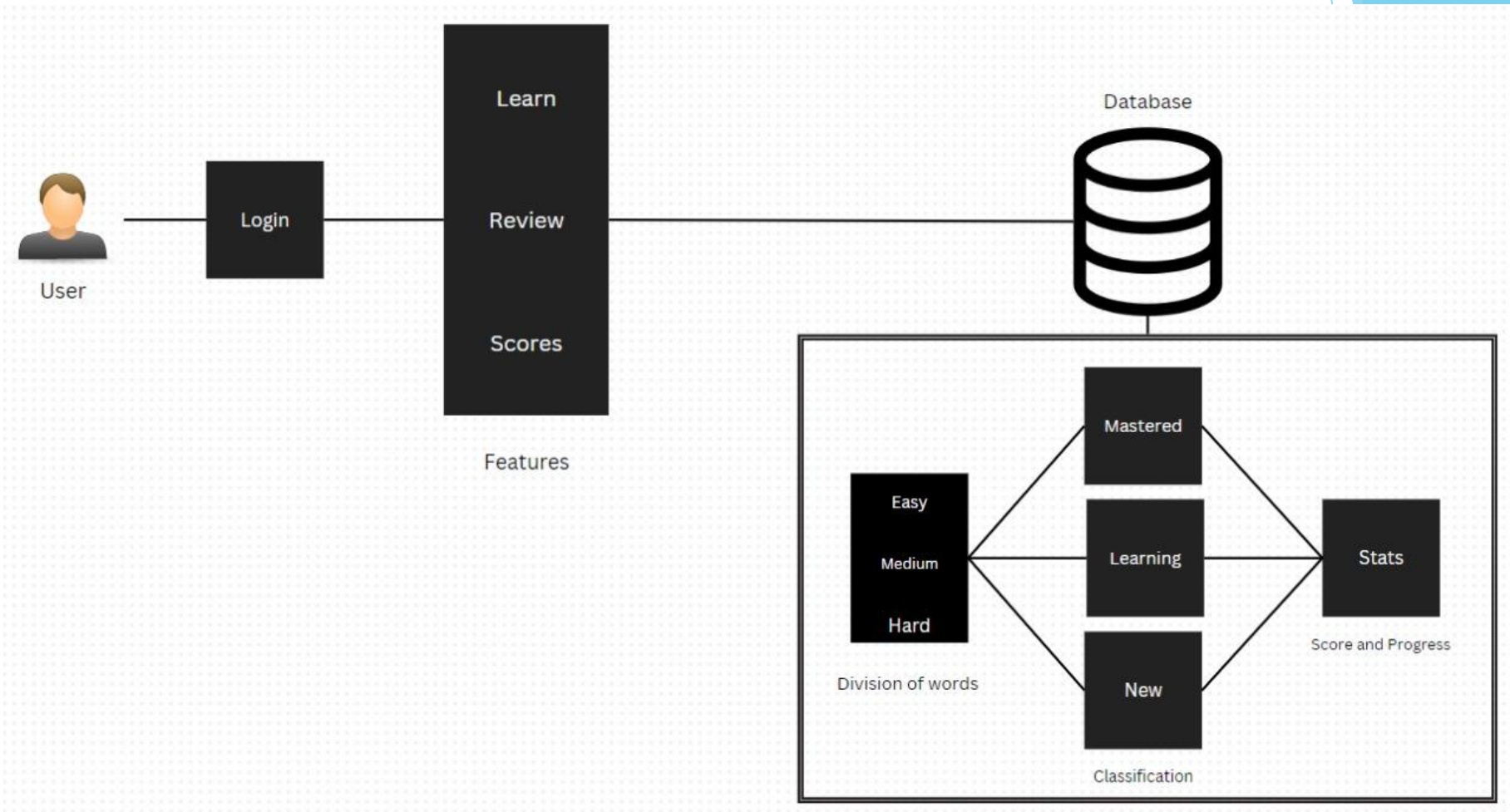
Limitations of existing systems

- Apps fail to accurately change the difficulty of the words leading the students to lose interest to continue learning.
- Most of the apps concentrate on a huge dataset at one time whereas Vocablearn divides it into learning sets which makes it easier to learn and log the learnt words.
- Words are not accurately sourced from valid datasets which causes a learning gap and reduction in quality of learning.
- Websites fail to make the UI interactive and interesting which takes away the excitement from the learners.

Problem statement

- Learners often face challenges in systematically expanding their vocabulary and retaining new words over time.
- Traditional vocabulary learning methods lack personalized revision strategies, fail to adapt to individual learning speeds, and to ensure sustained progress. As a result, learners struggle with inefficient memorization, inconsistent practice, and a lack of measurable improvement.
- Learners often do not have the tools to efficiently review previously learned words, leading to rapid forgetting.
- Without measurable outcomes or feedback loops, learners lack the motivation to continue, as they cannot clearly see how their vocabulary is improving.

System Design



Technologies and methodologies

- **Front-End :**

1. HTML
2. CSS
3. JS (Javascript)
4. Python

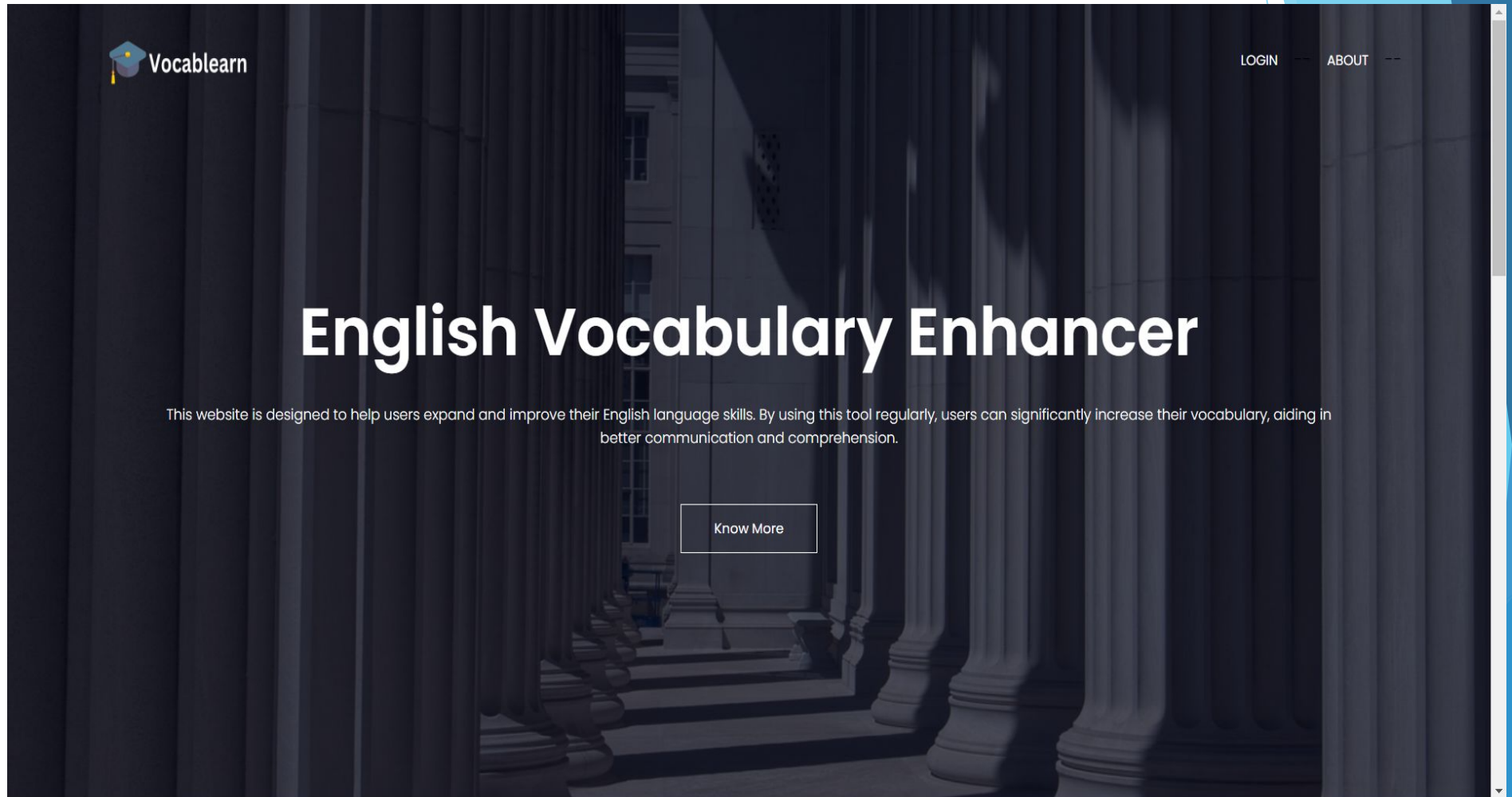
- **Back-End :**

1. MySQL
2. Flask

- **Algorithms :**

1. Leitner system
2. Decision tree classifier
3. Linear Regression

Implementation



Implementation

Why Choose Vocablearn ?

Vocablearn has various features for improved and better user experience.

Word Review



The user can go through the words that they have mastered, hence can be revised.

Score Viewing



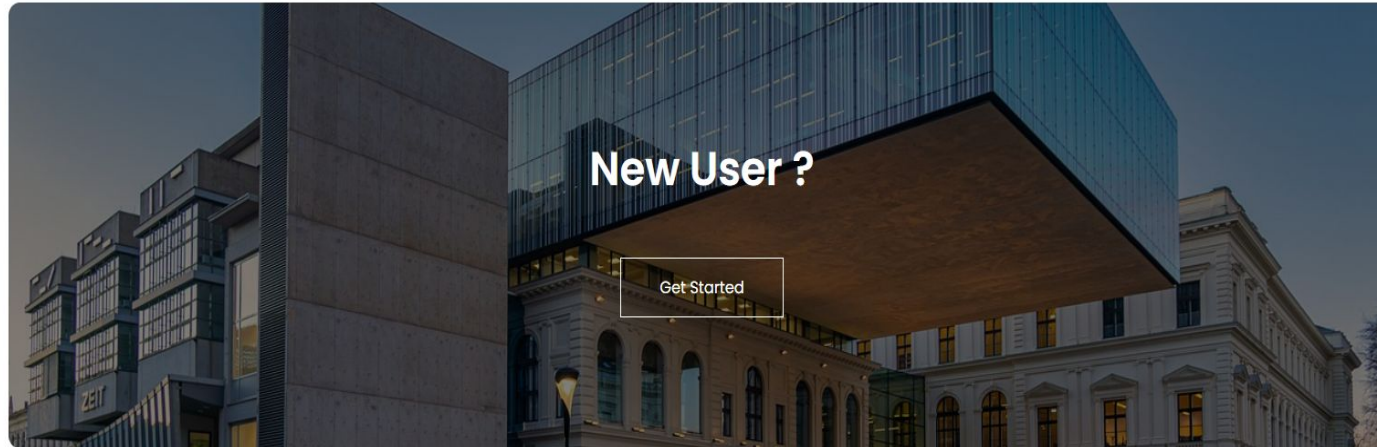
Based upon the progress and performance of the user, scores will be allotted and can be viewed in the form of graphs.

Auto Word Adjustment



According to the performance of the user, the difficulty level is altered by the model automatically.

Implementation

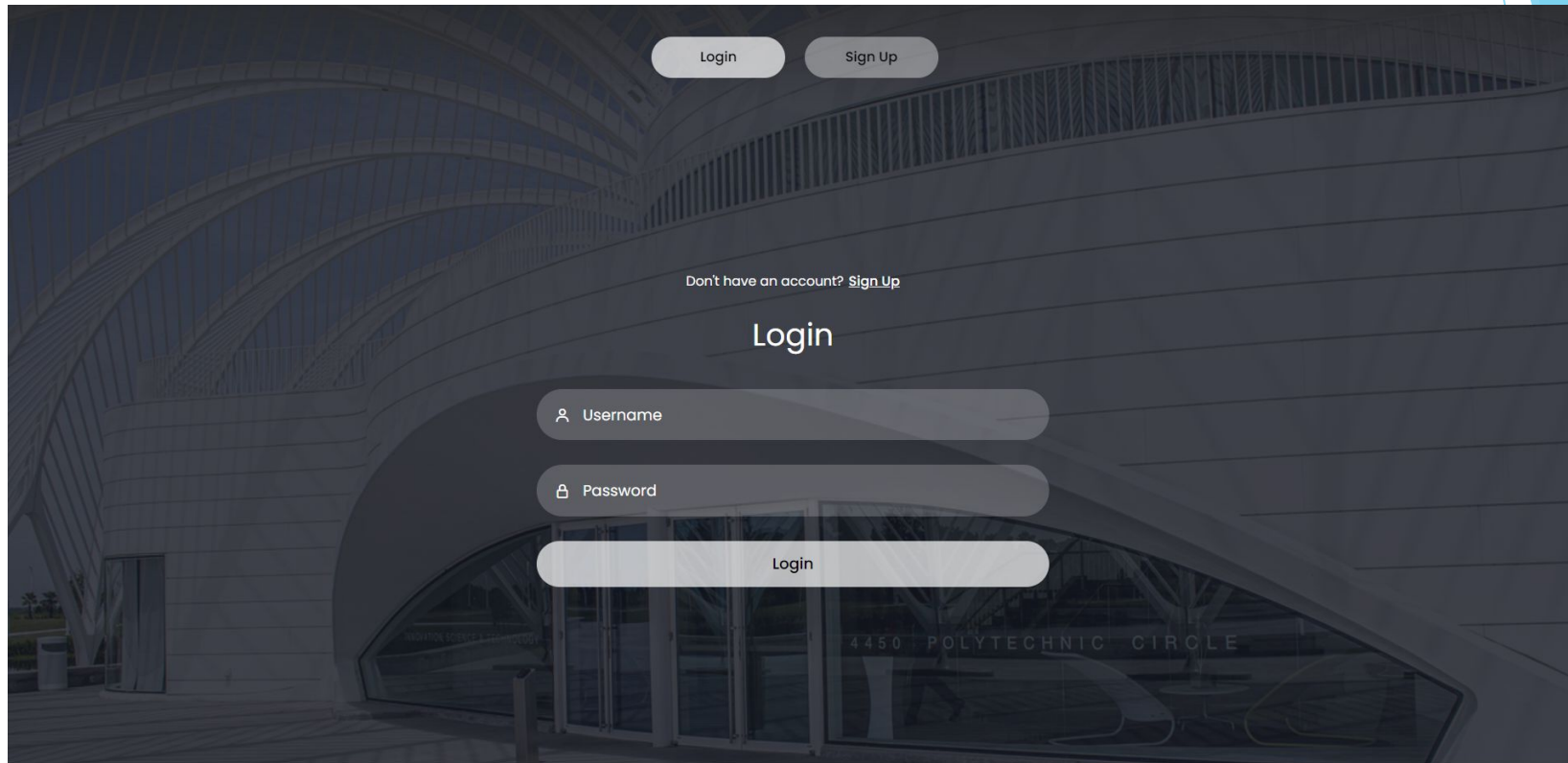


About Us

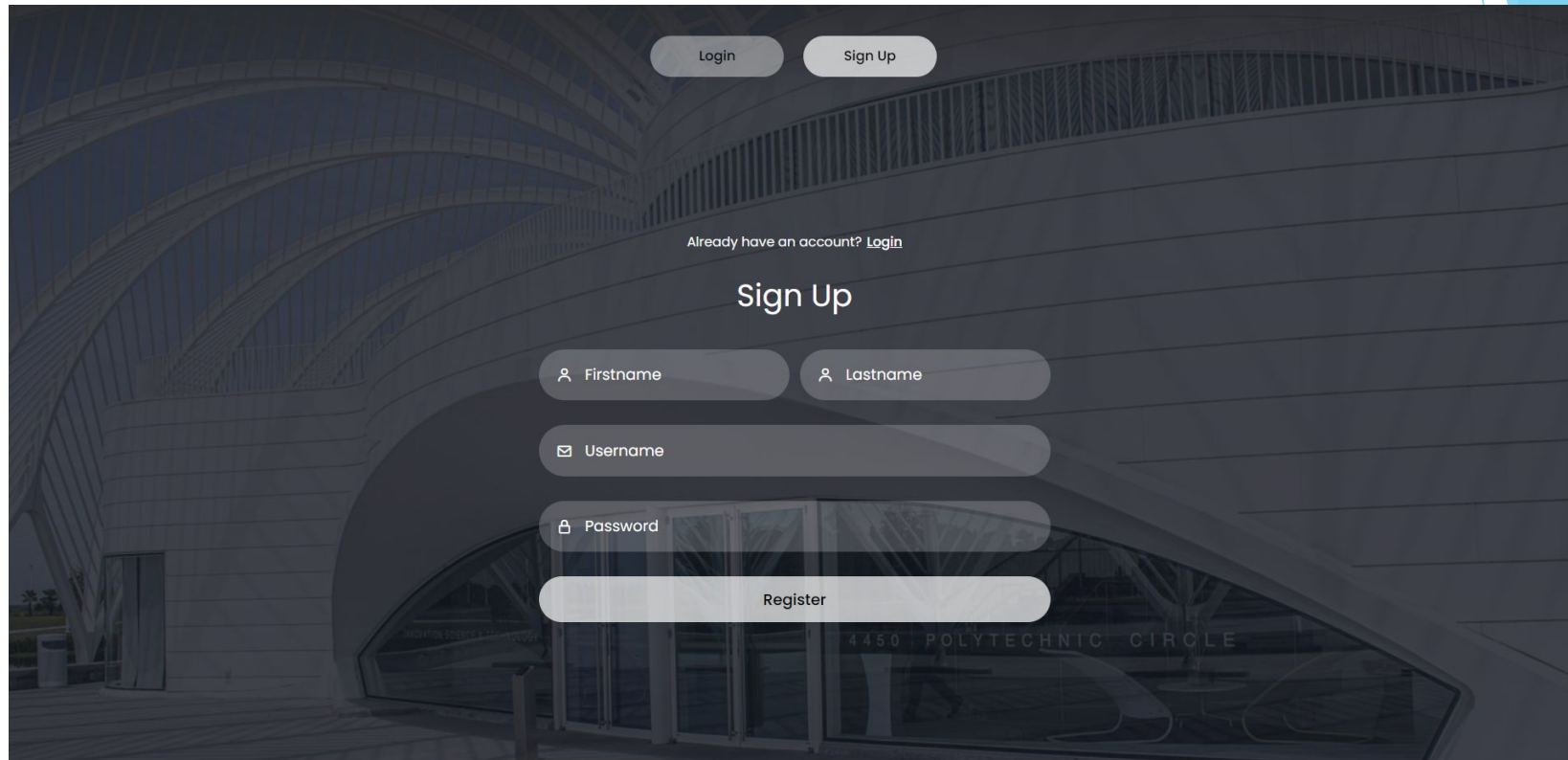
gaurisasikumar04@gmail.com
A. P. Shah Institute Of Technology



Implementation



Implementation



Login Sign Up

Already have an account? [Login](#)

Sign Up

Firstname Lastname

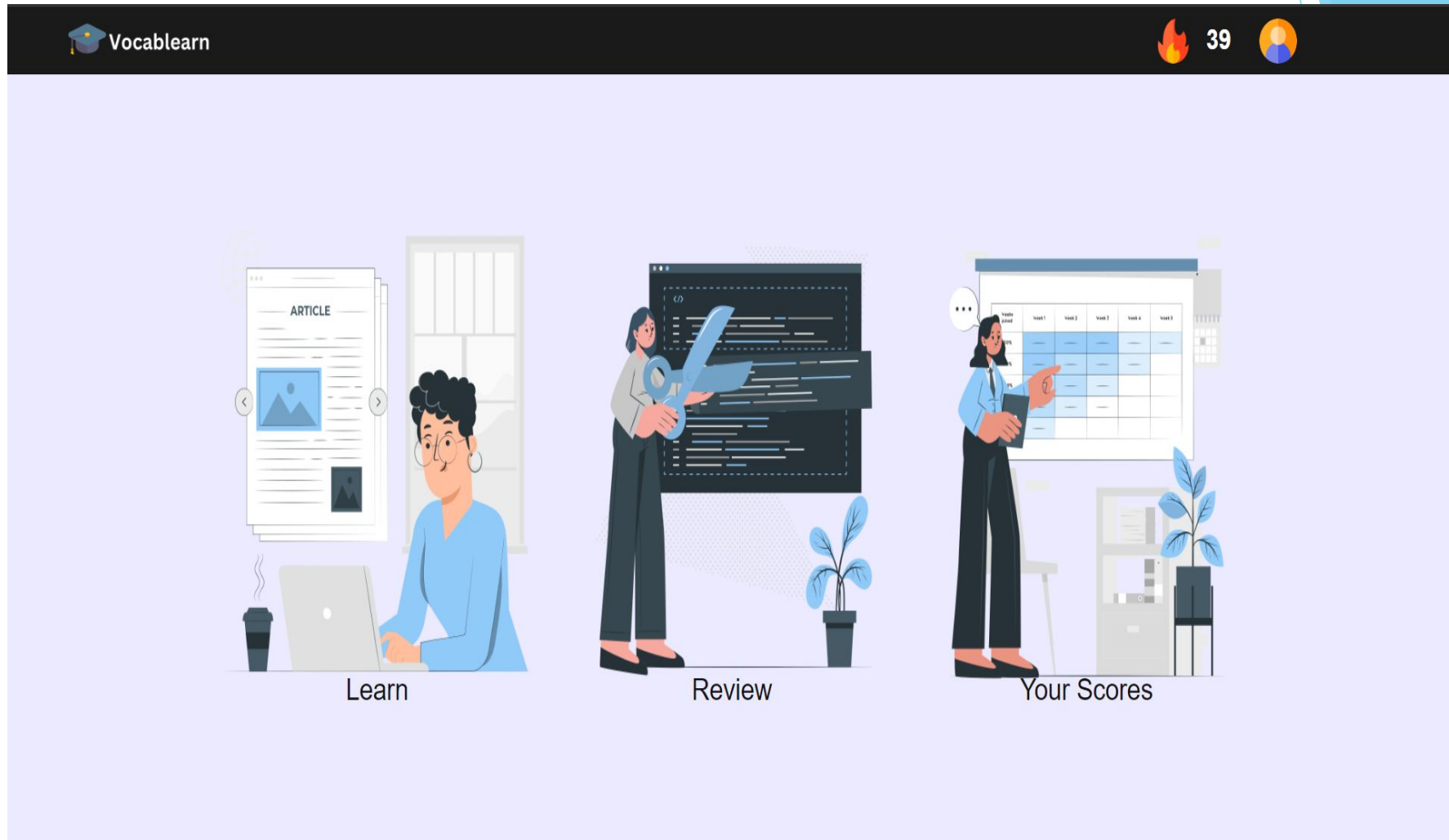
Username

Password

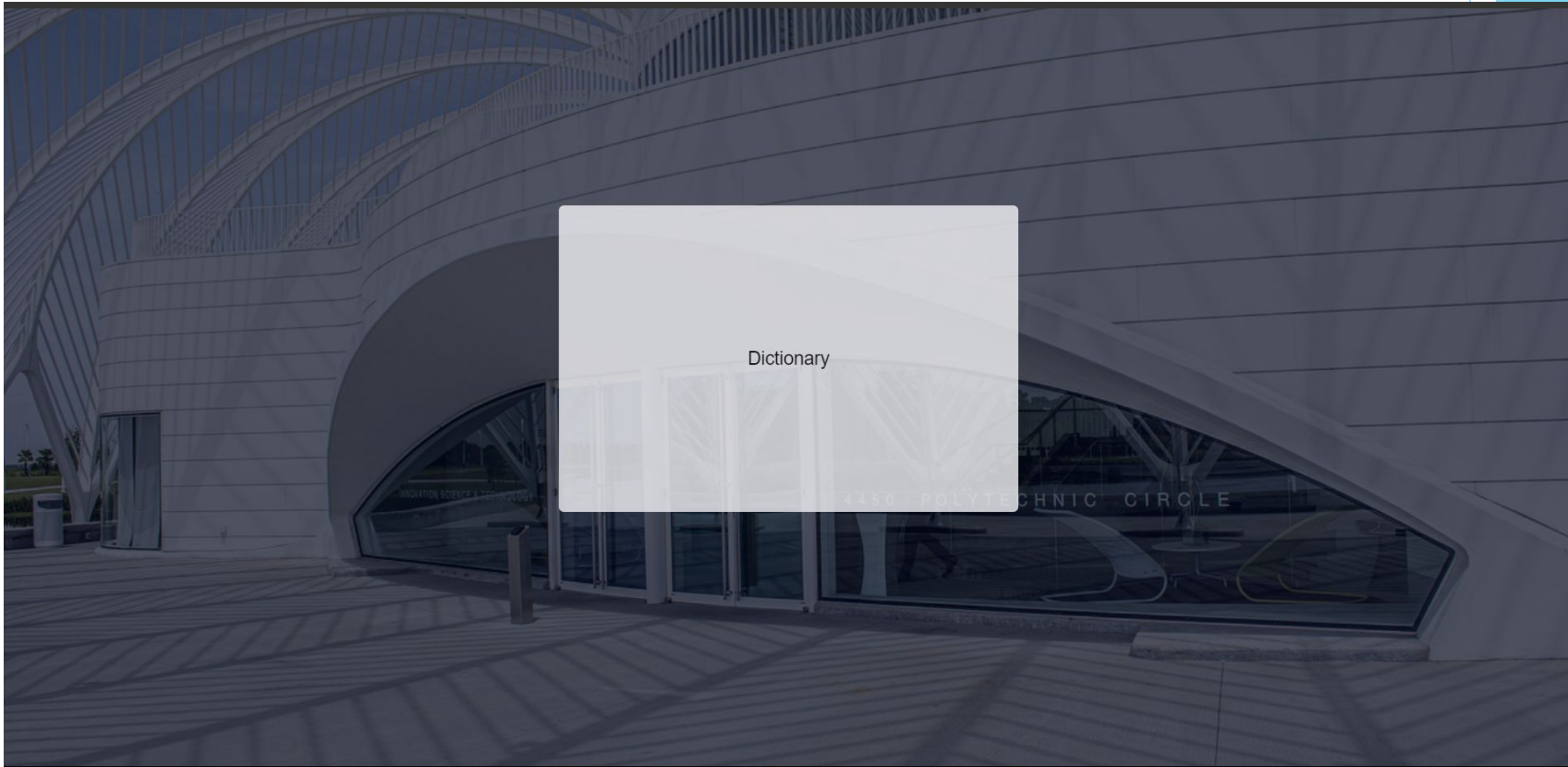
Register

4450 POLYTECHNIC CIRCLE

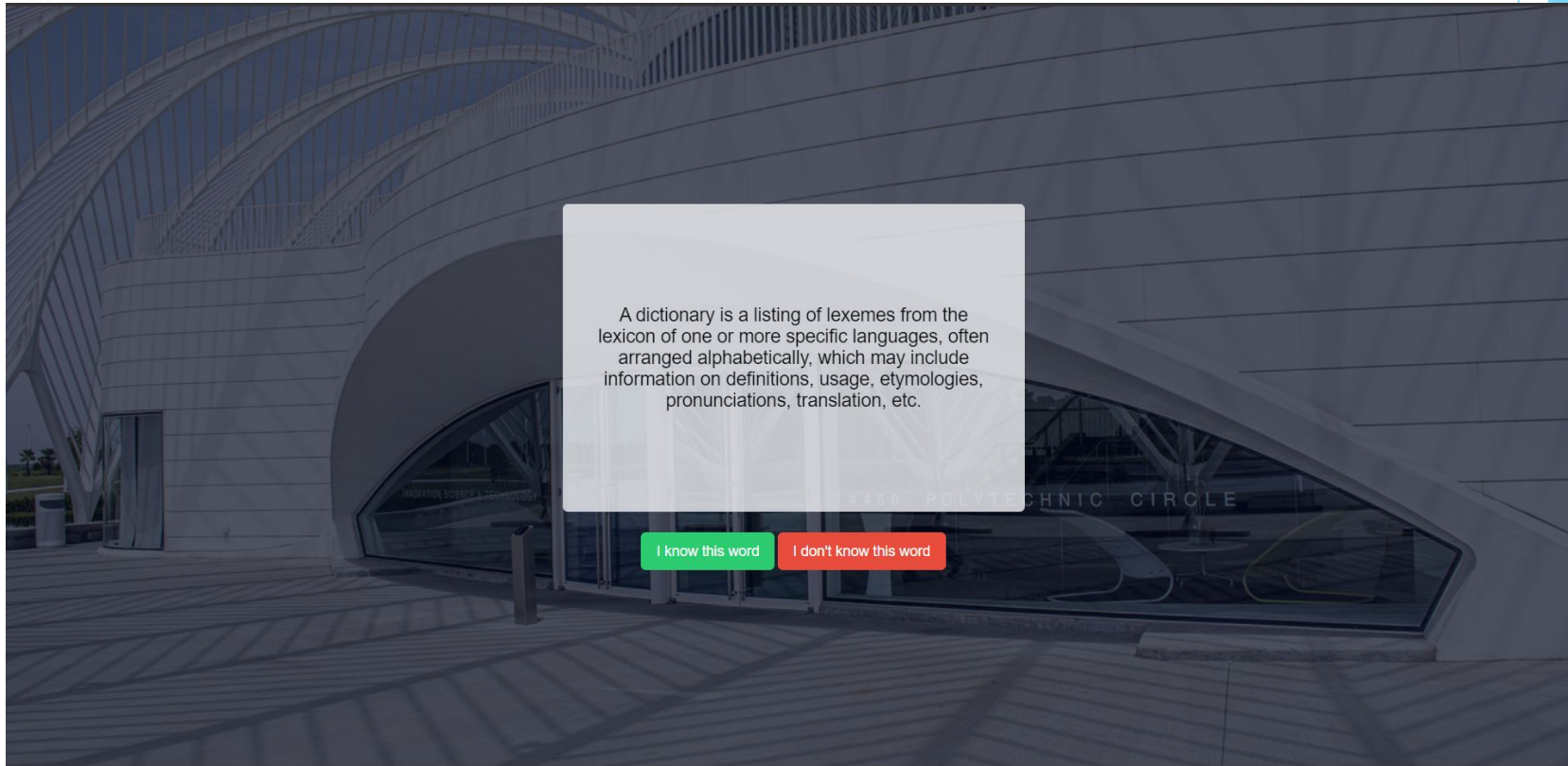
Implementation



Implementation



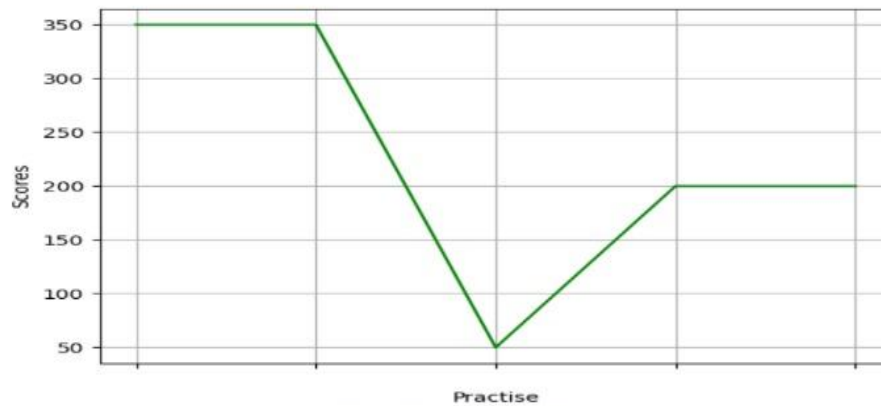
Implementation



Implementation

Words Mastered: 3
Total Words: 370

Learning Words: 34
Total Words: 370



Conclusion

- "Vocablearn" is designed to make vocabulary learning easier and more effective by focusing on what learners really need: a personalized, engaging experience.
- It helps users overcome common struggles like forgetting words, inconsistent practice, and a lack of motivation.
- By offering tailored revision sessions, tracking progress, the platform makes learning more structured and enjoyable.
- Whether you're preparing for exams or just looking to improve your language skills, "Vocablearn" aims making vocabulary building a rewarding experience.

References

[1] Miao Peng , Yujie Wu , Yannan Qiu, “English Word Difficulty Classifier Based on Random Forest Model” , Academic Journal of Computing & Information Science, 2023 francis-press.com, 2023

[2] Xilun Li , Lingyu Sun , Jigang Li, “Word difficulty classification prediction based on ARIMA and DBSCAN” , School of Management Science and Engineering, Dongbei University of Finance and Economics Dalian, China, IPIIS 2023

[3] Shivam Parihara, Shubhi Miradwal , Abhinav Panse , Ruchi Patel , “Word Difficulty Level Prediction System Using Deep Learning Approach”, Computer Science Engineering Department, Medi-Caps University, Indore, India, Topics in Intelligent Computing and Industry Design (ICID) 2(2) (2020) 109-112

References

[1] Dr. P. Santhosh, Mr. M. Rajesh, “*Enhancing the Vocabulary Learning Through Android Applications*”, International Journal of Advances in Engineering and Management (IJAEM), Volume 4, Jan 2022

[Enhancing the Vocabulary Learning Through Android Applications](#)

[2] Farhana Diana Deris, Nor Seha A Shukor Universiti Teknologi Malaysia, Sukadi Johor, Malaysia, “*Vocabulary Learning Through Mobile Apps: A Phenomenological Inquiry of Student Acceptance and Desired Apps Features*”, Universiti Teknologi Malaysia, 2019

<https://www.learntechlib.org/p/216526/>

[3] Jalil Fathi, Forough Alipour, Abdulbaset Saeedian , “*Enhancing Vocabulary Learning and Self-Regulation via a Mobile Application: An Investigation of the Memrise App*”, Journal of Modern Research in English Language Studies, 2018

https://jmrels.journals.ikiu.ac.ir/article_1640.html

Thank You !!