



Real-Time Qur'an Error Detection Using Machine Learning and Natural Language Processing

Manar alsaedi, Azzah mahjoob, Amira alotaibi, Waad alsaedi, Atheer Alsaedi

Supervised by: Dr. Ashwag Omar Maghraby

Project ID: UQU-CS-2021F-11

Dept. of Computer Science, Umm Al-Qura University, KSA, 2022.



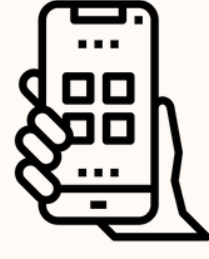
Abstract

Holy Qur'an recitation is imperative to worship and actions of reward for Muslims worldwide. It is also the duty of every Muslim to read the Qur'an correctly in classical Arabic. However, many people fall into mistakes when reading the Holy Qur'an, and those mistakes are forbidden. Mistakes can be missing or replacement words, verses, or misreading Harkat or Tajweed. This project proposed a solution to recitation and learning Holy Qur'an quickly via using a machine learning approach and natural language processing.

Related Works



Al-Moallem



E-Hafiz



Tasmee

There are some problems in these applications:

- (1) Not detecting all types of mistakes.
- (2) Failing to clarify mistake types.
- (3) Slow response.
- (4) Non-real-time application.

Objectives

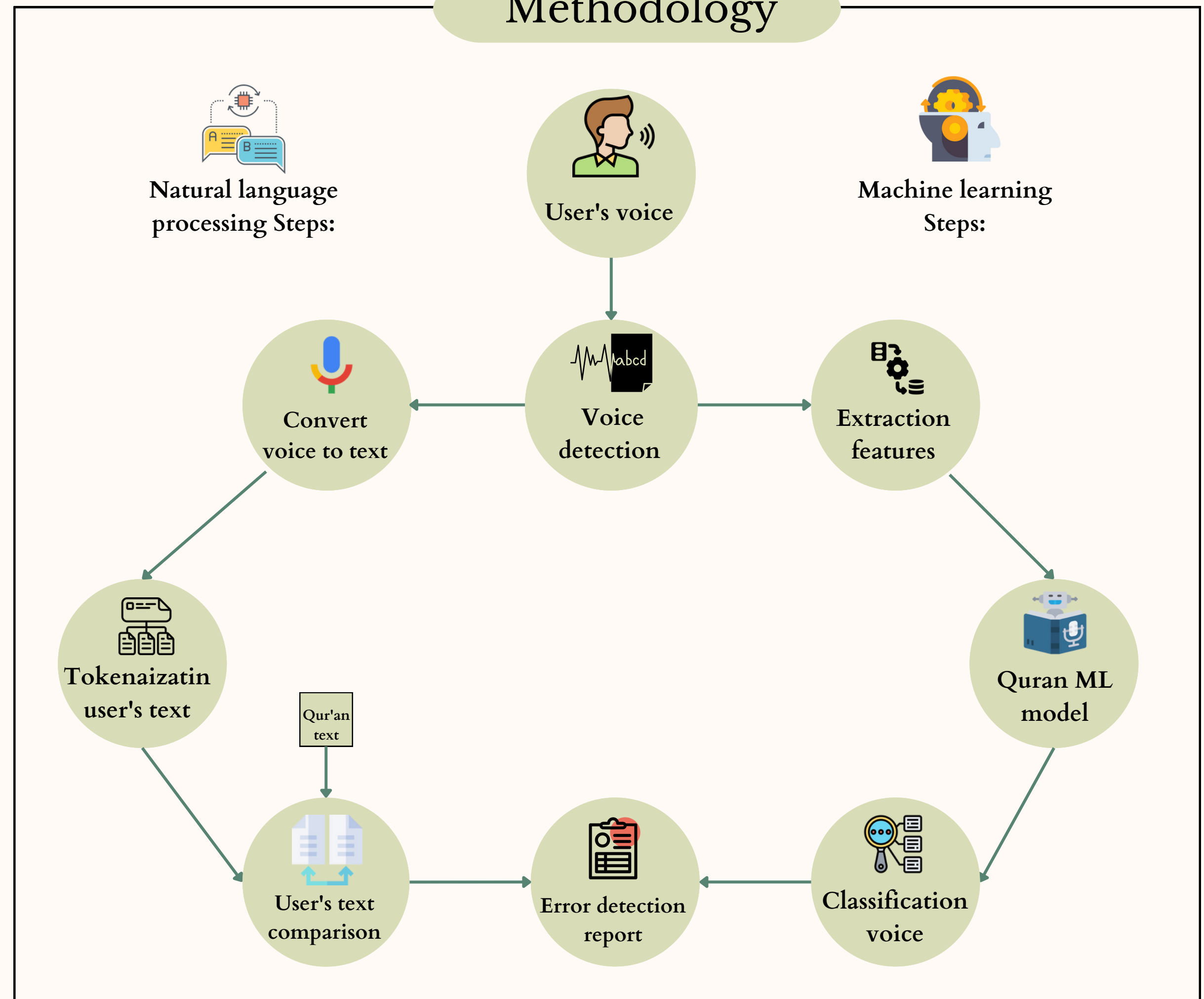
There are three main objectives of this project:

Collect a set of records of Surat Al-Ikhlās, build a dataset, and prepare it for the task of detecting people's Quran recitation to determine whether their recitation is correct or not.

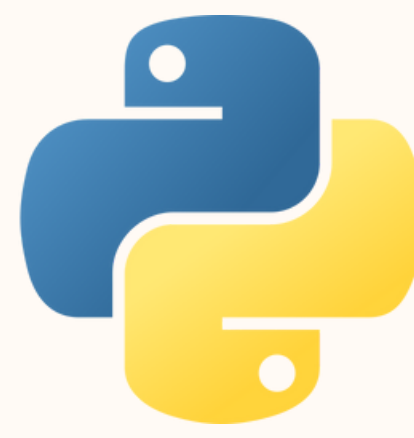
Create a machine learning model by trying different classification algorithms and choose the best model with the highest accuracy, to detect and correct most common recitation (Harakat and Tajweed) mistakes.

Use Natural language processing to detect and correct most common memorizing (Missing, Replacement, Addition, and Repetition) mistakes.

Methodology



Development Tools



Python language



Jupyter notebook



Spyder IDE

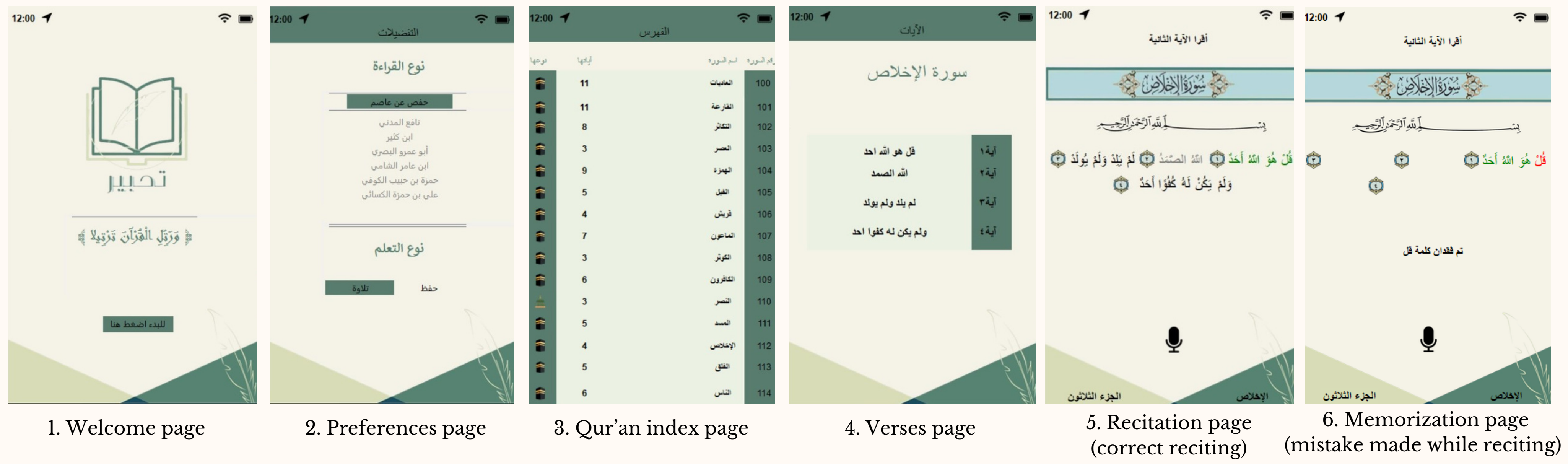


Google speech API



Microsoft excel

User interfaces



Conclusion and Results

This research ends with a real-time application called "Tahbeer" that works to detect and correct the most common mistakes in reciting and memorizing of Surat Al-Ikhlās using both machine learning models and natural language processing. The dataset includes 1506 records, were collected utilizing Google forms and Qur'an audio websites, and then labeled manually. The result showed that Random Forest achieved the highest accuracy with 64%, and the real testing results, on 70 persons, of the Qur'an using the Tahbeer system, are very encouraging where the application detected 59% of the mistakes.

Future works

Some different improvements have been left for future work, such as:

- (1) Increase the number of records for the dataset and gathered them from non-Arabic speakers.
- (2) Expand the test to more than 70 persons to increase the accuracy of the application.
- (3) Extend the application to cover all Surahs and Qira'a types.

Group email: Tahbeerapp@gmail.com

Contacts: Twitter: @TahbeerApp

Supervisor email: aomaghraby@uqu.edu.sa

References:

