Expressify

Aditya Khuntia

Divyaansh Seth

AMITY INTERNATIONAL SCHOOL

(Under the aegis of Ritnand Balved Education Foundation)
M Block, Saket, New Delhi - 110 017
Ph. No. 011 - 40738400, 40738421
E-mail ID : principal@aiss.amity.edu, Website : www.amity.edu

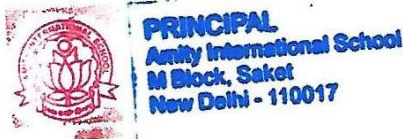AIS/CSIR/2020-21   18.04.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that the project proposal which is being submitted for the competition for CSIR Innovation Award for school children 2021 is an authentic original creative work of the students.

Divya Bhatia
Principal

PRINCIPAL
Amity International School
M Block, Saket
New Delhi - 110017

QUALITY SYSTEM CERTIFICATION
DNV·GL
ISO 9001:2015

AMITY INTERNATIONAL SCHOOL
(Under the aegis of Ritnand Balved Education Foundation)
M Block, Saket, New Delhi - 110 017
Ph. No. 011 - 40738400, 40738421
E-mail ID : principal@aiss.amity.edu, Website : www.amity.edu

AIS/CSIR/2020-21   18.04.2021

TO WHOMSOEVER IT MAY CONCERN

This is to certify that the details of participants and the title of the project which has been submitted to CSIR are as follows:

Title of the Innovation - Expressify

Participant Name -   Aditya Khuntia
Date of birth -   12th October 2005
School Address -   Amity International School, M Block, Saket
Residential address -   Flat No: 1029, C-1, VasantKunj,New Delhi -110070
Class -   10-C
Phone: 981851259
E-mail address -   adityakhuntia5@gmail.com

Participant Name -   Divyaansh Seth
Date of Birth -   25th August 2006
School Address -   Amity International School, M Block, Saket
Residential address -   92- Bakhtawar Singh Block, Asian Games Village,
                        New Delhi - 110049
Class -   10-D
Phone: 8851572246
E mail id -   seth.divyaansh@gmail.com

Thanks & regards

Divya Bhatia

Divya Bhatia
Principal

PRINCIPAL
Amity International School
M Block, Saket
New Delhi - 110017

DNV·GL
ISO 9001:2015

Council of Scientific and Industrial Research Innovation Award for School Children
2021

Expressify, creating a Software Application to solve Societal Problems relating to Communication Skills of People with special emphasis on Global Phrases, Idioms, Culture and Regional Heritage

---

	Aditya Khuntia	Divyaansh Seth
Date of Birth:	12th October 2005	25th August 2006
School:	Amity International School, Saket	Amity International School, Saket
Class:	10-C	10 - D
Residential Address:	Flat No: 1029, C-1, Vasantkunj, New Delhi - 110070	92- Bakhtawar Singh Block, Asian Games Village, New Delhi - 110049
Telephone Number:	+91 9818851259	+91 8851572246
Email Address:	adityakhuntia5@gmail.com	seth.divyaansh@gmail.com

Our prototype is available at https://eytoy2.pythonanywhere.com

---

Abstract: *Creating an interactive design-based application to provide a platform for all age groups to improve communication and linguistic skills in a practical manner with help of specific scenarios designed to diversify their negotiation, manipulation and general conversation skills specifically relating to Idioms, Global Phrases, Regional Culture and Heritage and solving the societal problems of lack of communication skills for Non-Native Speakers all across the world.*

THE PROBLEM

In today's global interconnected world <u>English has become a universal language</u> like a wide banyan tree housing amongst its branches many races and religions. For <u>*Non-Native Speakers*</u> (i.e., Speakers for whom English is a foreign language), understanding English of <u>*Native Speakers*</u> (Speakers for whom English is a primary language) and expressing themselves has become quite difficult. The following problems have become quite evident nowadays:

➢ Feeling of exclusion while interacting with a group of Native Speakers mainly due to <u>lack of understanding of certain specific vocabulary during the conversation.</u>

➢ <u>Confusing interaction with Native Speakers due to their usage of Idioms and Phrases</u> such as "*mea culpa*", "*Bite off More than you can chew*", "*hot potato*" etc.

➢ Not being able to precisely converse with native speakers about your own culture and heritage due to the <u>presence of many stereotypes and cliches about their culture</u>.

➢ Non-Native Speakers being <u>afraid about the pronunciation of certain words and phrases due to having an accent</u>

➢ Being <u>perceived as naive or clueless due to lack of understanding and reaction</u> of Non-Native speakers while conversing with the Native Speakers.

For e.g. The famous idiom "to put lipstick on a pig" (i.e., making superficial changes to something to hide its fundamental failures) is used in a conversation, a <u>person with less knowledge of the idiom might take it literally as a challenge of sorts and might also laugh on the funny choice of words.</u> This misunderstanding would result in a completely different reaction from what was expected of the person thus making them look naive or clueless

"Why are you putting wallpaper up when the entire farmhouse is falling apart around you? You are just putting lipstick on a pig."

➢ There is a need to <u>keep a broad vision for understanding of English phrases.</u> With England ruling more than half the globe at one time, the adoption of English as the local language was unavoidable. Resultantly, new variations of English emerged in the countries that were under the British dominion.

<u>Expressions combining elements from English with indigenous forms</u>

Nigerian English t*o put sand in someone's gari* ('gari' is a type of flour) which means 'to threaten' someone's livelihood;

<u>Translations of idioms from indigenous languages.</u>

"to shake legs" mean*s* to be idle. It is used in Singapore and Malaysia prominently as a translation of the Malay idiom *'goyang kaki'* the blending of two existing idioms giving way to a new idiom

<u>Thus, many phrases are not original to colloquial English but instead are direct translations and borrowed from different languages and culture. To understand them would always be a difficulty</u>

A difference in native language leads to a person not being able to fully convey their point in a situation with people of different language backgrounds which can lead to miscommunication. What <u>may seem like a good topic of conversation to one might seem rude to another</u>. For e.g. For Farsi speakers, Hindi, and Urdu speakers, the history of flowing, poetic, grandiose descriptions take the day. So just saying plain facts one after the other is considered almost rude and so conversation spins around many made up stories.

As long as languages are confined to their territorial borders, there is no big problem while using traditional cultural expressions. But as humanity got more interconnected through the World Wide Web, languages started crisscrossing and such small phrases started creating short circuits in communication. So, it is very important that there are steps taken to ensure that our future generations are successfully able to overcome these challenges and have better and clearer understanding and vision.

THE SOLUTION

Design (GUI)

For the application, we created a user friendly and attractive design which can retain the attention of children for long periods of time. Our design was made keeping in mind the interactiveness of the application and all the bots given in conversations were given unique identities that fit the pace and style of conversation.

Furthermore, our design was created with simplicity in mind and is created in a way that it can be used without any prior instructions and can be navigated through easily. While retaining our simplicity, we also created the application to be very attractive and eye catching. This was done through features such as the phrase gallery in which we added hover effects, brighter color scheme and easy to navigate options. Furthermore, an instruction page is also present, to help the users in case of any confusion or discrepancy.

Our Logo:

For our logo, we chose something similar to the iconic arcade video game character Pac-man. Pac-man was chosen as the base for the design as it is a character that does not speak at all but expresses himself very clearly. Furthermore, Pac-man is a very relatable figure as it was a childhood favorite for most and therefore was the best way to represent our application


Novel Features

Pronunciation Check:

In Our application we have included a pronunciation check, wherein the user is provided with a random Global phrase and is required to pronounce it. Using the Speech Recognition Technology, we compare the output to that of the word, if they match the user has pronounced the word correctly, else the user can try again.

Since Global Phrases come for different regions and have different dialects (For example Au Courant - French; Ipso Facto or Persona Non Grata - Latin; Nanty Narking - Victorian English), we have included both a regional pronunciation and a general English

pronunciation for the user to speak. If required the user may go to the phrase gallery and hear the pronunciation

Using a NLP (Natural Language Processing) Processor we provide a confidence rate in the output to showcase how well numerically has the user pronounced the phrase. Anything above a 0.75 is considered good and correct. Thus, our app perfects the user's pronunciation as well in a very simplistic way.
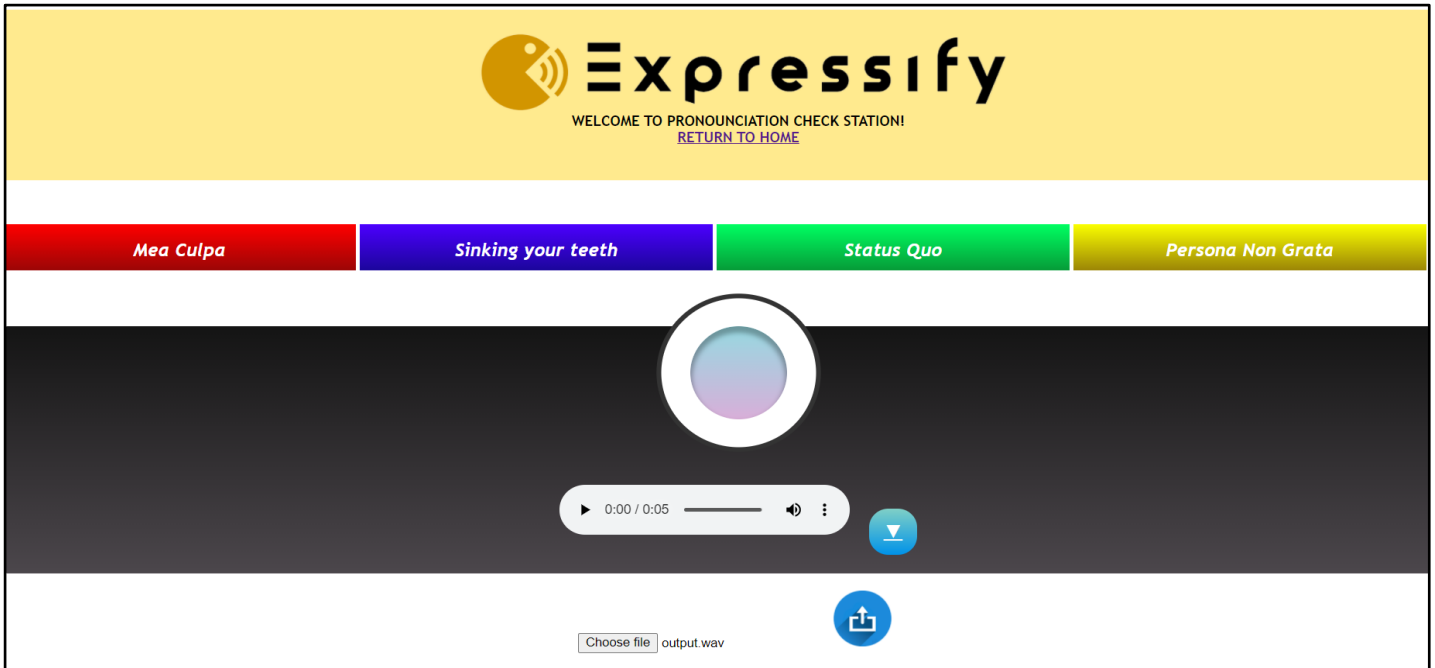
Expressify Pronunciation Check Page
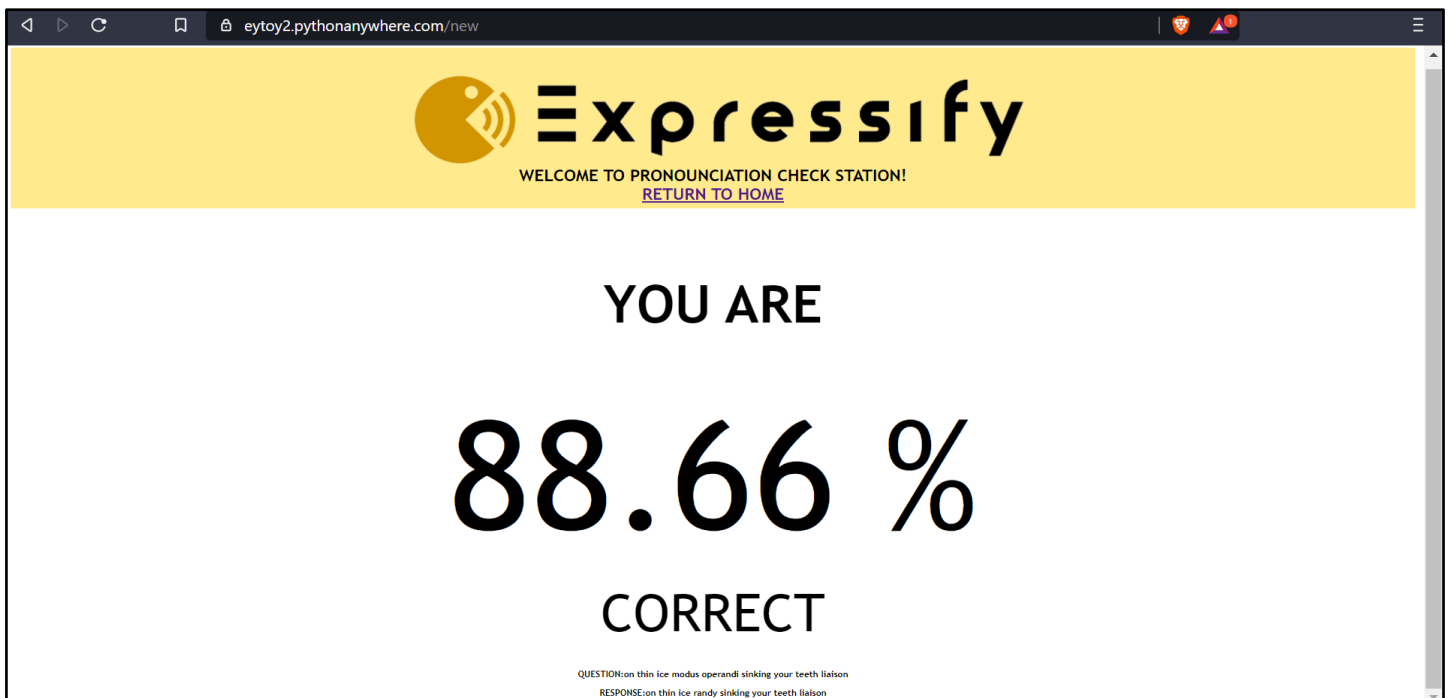
Expressify Pronunciation Result Page

<u>Neural Network:</u>

We used Sentimental Analysis in combination with a neural network to automatically assess the user's usage of idioms. We had to first create a proper distinction between general phrases and Idioms, Global Phrase. *"2+ word expressions (phrases made up of at least 2 words).However they are different as their meaning cannot be guessed from the literal meaning of each word in the phrase*

We organized 500 idioms in 16 categories which can directly such as (happy, sad) or indirectly such as (success, failure) be corresponded with human emotion and general daily usage.

Theme	Total	Theme	Total
Anger/Annoyance	45	Mistakes/Errors	5
Anxiety/Fear	14	Politeness	8
Arguments/Disagreements	37	Problems/Difficulties	57
Enthusiasm/Motivation	10	Safety/Danger	27
Feelings/Emotions	48	Sleep/Tiredness	11
Fun/Enjoyment	22	Success/Failure	84
Happiness/Sadness	21	Surprise/Disbelief	16
Madness/Insanity	11	Violence	6

Since there does not exist any readymade annotated dataset, we used 3 annotators to manually tag whether the idiom is of positive/negative or neutral sense. Similarly, we used 5 annotators to tag the idiom when used in a specific context. An idiom was considered to be universally true with its respective emotion when more than 50% annotators tagged it as the respective emotion.

"All right, do not *jump down my throat*" was annotated as negative by the first two annotators but positive by the third one, the ground truth associated with the sentence was determined as being negative.

We used a Lexico Syntactic Pattern (string matching based on tokenized value and synaptic structure) to recognize idioms in a text. Since verb, tenses may change in an idiom even though its structure remains the same, thus string matching would not have sufficed.

After this, we used a machine learning tool (keras) in python to train the classifier model and got an accuracy rate of 85.5%. We have now incorporated this model in our model conversations to give specific marks to every option in the positivity meter (Measures the tone of the conversation), and have further used them for idiom usage correction in the practice mode from user's input.

Realistic scenarios:

Our application contains well researched scenarios which are based on objective scenarios which we ourselves along with some people we surveyed had encountered in real life. These scenarios help the user in getting straight to the point in a conversation as that is how most real-life conversations take place. This can also help the user learn in a much easier manner.


This is an image which is shown to the users at the end of their conversation which shows them the possible outcomes they could have gotten through various responses.

Phrase Gallery:

Our application includes a phrase gallery containing all the phrases in our large dataset that have been used in model conversations practice throughout the app. We have provided it's respective contextual usage as well as audio pronunciation in the application. In our proof-of-concept study, this was the 2nd most effective way of learning more about idioms and thus, we incorporated it by creating an eye-catching page.
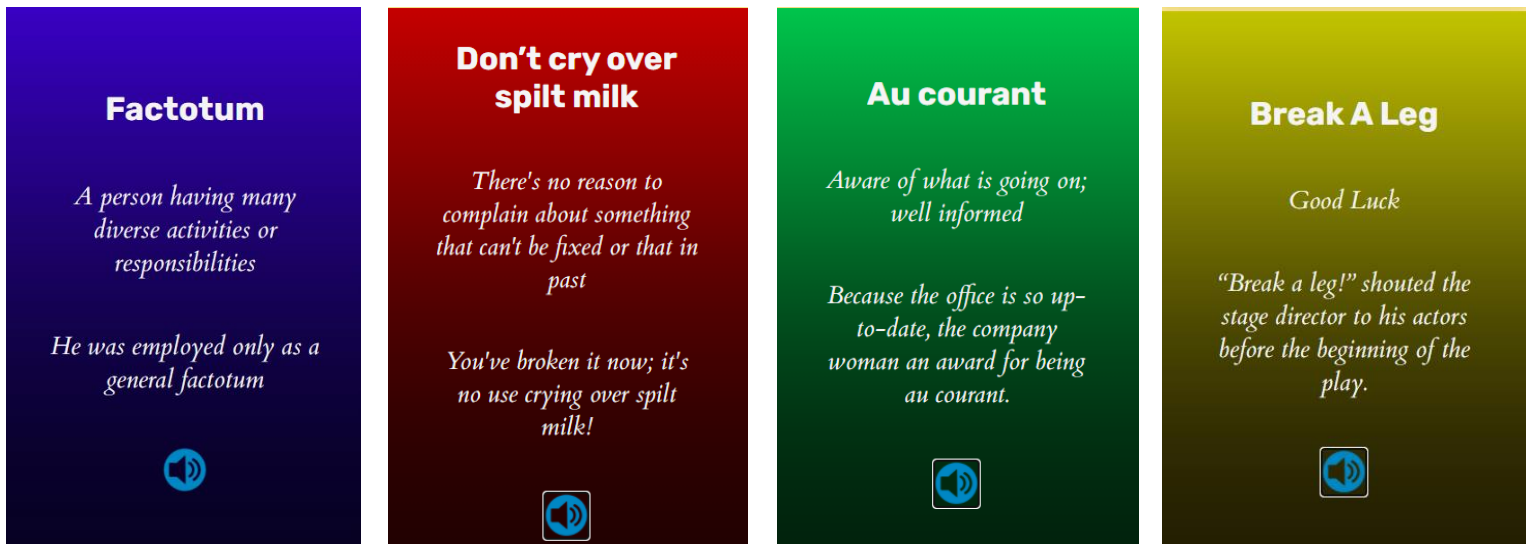
Expressify Phrase Gallery Cards
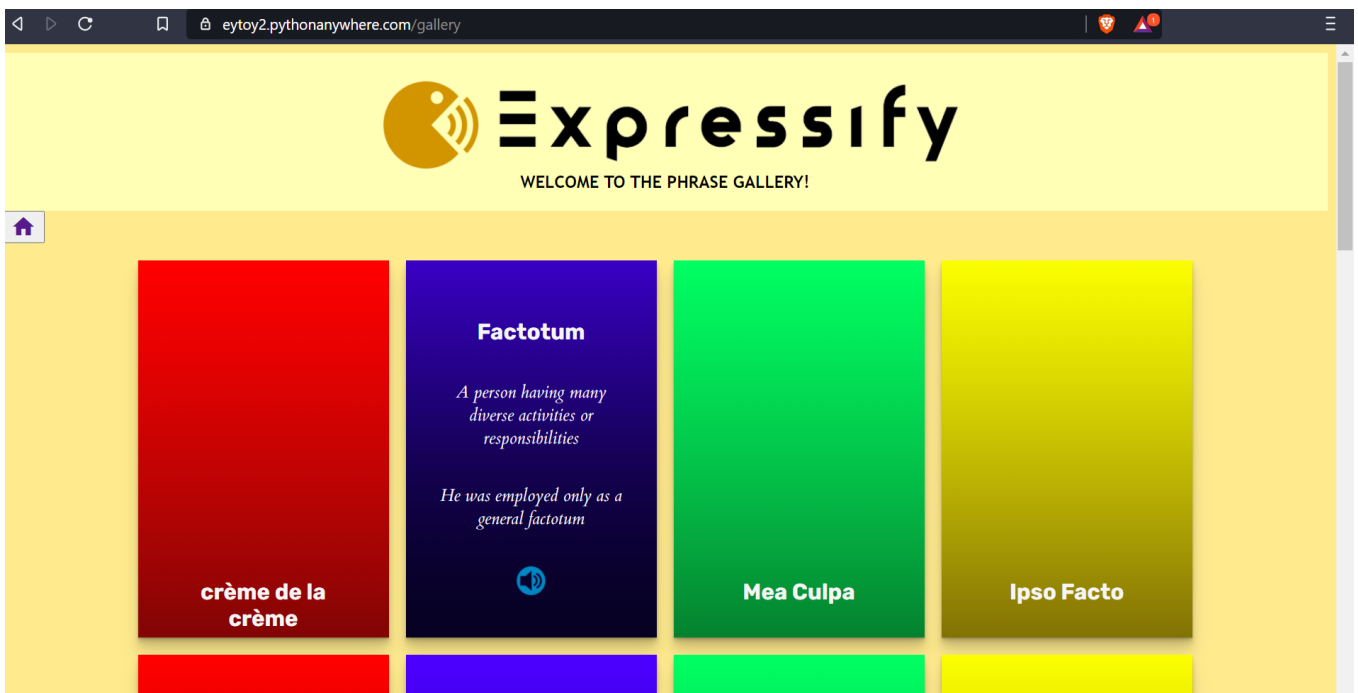
Expressify Phrase Gallery Page

Utility (Specific Scenarios)

Our application has a very wide scope and variety for different people belonging to different professions and ages in terms of its usage. A few of them are mentioned below.

<u>For school students (Primary to High School):</u>

School students can use this program to improve their English usage skills which are highly. useful when interacting with students from other countries (Both Native speaking and Non-Native Speaking countries) so as to create an efficient medium of exchange in places such as competitions, meets, collaborations etc. at the global level. For students, the model conversation practice mode has been found most useful in our proof-of-concept study, using which they gain an all-round knowledge about idioms in respect to its Usage/pronunciation and Meaning.

<u>For college students:</u>

For students who aspire to go to or are already in foreign universities, it can sometimes be hard to adjust to the new and different ways in which people there may speak and understand English. Communication among their peers becomes difficult and almost impossible to give any reaction to whatever one might say. Therefore, this application can help college students by training them to refine their English skills and mannerisms which are used in English in that region.
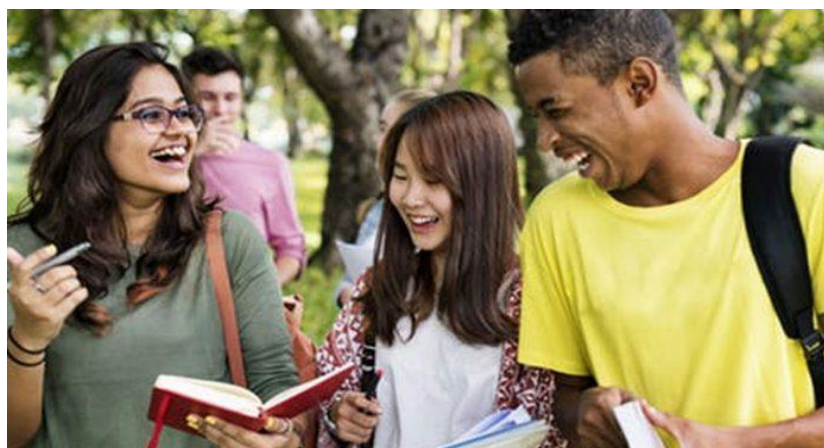
For professionals:

For people working in MNC's or in jobs that may lead to them interacting with foreign professionals on a regular basis, it can sometimes be hard to comprehend the other person or get your own point across due to a difference in mannerisms and specifics of spoken English. Furthermore, the issues of differing accents may also be very prevalent. Thus, professionals can use our application to improve their skills in idiom usage in specific non-native English-speaking countries as they affect the global scale largely as well. The Application provides them with the pronunciation and contextual usage of not only idioms but also Global Phrases, something that is encountered a lot in Multinational Companies and Internationally in conventions too.

For teachers:

For school teachers, our application can come in very handy as it can be used as a tool to teach students about the correct usages of idioms in conversations as well as the differing English in different countries from a very young age. This can help the students to learn more and more as they grow and thus reduce the difficulties that these students usually face when applying to foreign universities, foreign jobs etc.

➢ Gives people especially the Non-Native speakers, the ability to converse with Native speakers without any hesitation and worries about being perceived as naive, as they will truly understand the context and usage of many global phrases and idioms.

➢ For Kids, they will develop an all-round understanding about phrases and idioms and will also understand their contextual usage especially through the model conversations.

➢ For native speakers, it also helps in fixing their pronunciations, especially in a language like English which also contains words and phrases or international dialect such as *Mea Culpa*.

➢ Through the model conversations and neural networks positivity meter response, users understand how to keep their conduct in real life scenarios, where they too have and objective and must converse in correspondence to it.

➢ Through our neural network users can also understand the tone each phrase and idioms are associated with, like lipstick on a pig is negative while go the extra mile is positive

➢ The global dialect in English, especially helps in expressing one's root heritage and culture for example (the English word "Juggernaut" is derived from the name of a deity of Odisha "Shri Jagganath") Such facts and phrases are always useful while conversing with an international audience.

PROOF OF CONCEPT

Before we created a solution, we had to see whether there truly exists distinctions on understanding and usage of different phrases day-to-day life while trying to speak about ones Regional Culture and Heritage etc., by native and non-native English speakers.

Participants

We included 60 people of different ages and professions so as to understand the distinction between usage of idioms by various people belonging to different professions, ages and regions and thus understand the main demographic (Based on the following table) affected by the problem.

MEAN AGE	45	21	13	44	28	15	7	15	19	33
OCCUPATION	MNC Executive	Alumni and Indian College Student (Foreign Univ) UG Sci	Indian School Student(English Medium)- 8th grade	Shop owner(10th level education)	English major PG Student (Foreign Univ)	School Student (English Medium)- 10th grade	Primary School Student	American High School Student	Amerigan College Student UG Sci	5th Grade Education Level Rural Resident

Instruments

To gather the data for the proof of concept, 3 data gathering instruments were used
➢ Vocabulary Level Test (based on idiomatic usage)
➢ A questionnaire regarding usage of idioms during conversation
➢ A semi structured meet with people participating to assess their pronunciation.

Idiomatic Proficiency test

An idiomatic proficiency test is a test to evaluate the present knowledge about idioms among various demographics of the population. This test was conducted as part of a simple vocabulary level test (VLT) which was conducted among all the participants.

Categorization of Idioms in Initial Study
➢ Phrases and Idioms which can be literally interpreted.
➢ Phrases and Idioms with Foreign Origin
➢ Phrases and Idioms which can be understood in isolated use.
➢ Phrases and Idioms which can only be understood when used in context

Design of Initial Idiomatic Proficiency Test
The initial idiomatic proficiency test (IPT) was designed to test the level of skill and proficiency of the participants in terms of pronunciation and knowledge about idioms. The test was set up in a way that when the users would start the test, they would be given an Idiom in the question and they would have to match it with its correct meaning and context.

Idiomatic Proficiency Test Sample Questions

We chose the idioms through a careful scanning process to find out the most used idioms mixed in with some idioms that are not as common. Furthermore, some idioms were included from non-native English-speaking countries to incorporate cultural and linguistic diversity of English across countries.

These idioms were then given labels from easy to hard based on the scores we received from the first test based on which categorization of idioms was done to problem specific areas for people belonging to respective age groups and professions.

Procedure of Idiomatic Proficiency Test

The Idiomatic Proficiency Test was distributed to the participants online. The Idiomatic Proficiency Test was introduced as a list of 40 questions without showing them the end of each level so they would try to answer as many questions as possible. It took the participants between 10 to 30 minutes to complete the Idiomatic Proficiency Test.

Bi-Weekly Meet

The bi-weekly meet is a further step on the way to gathering objective evidence on difficulties with learning idioms and the strategies used to understand idiomatic expressions. A major advantage of meeting online live is its adaptability and the direct response it can elicit from participants.

5 such sessions with 5 participants per session were conducted over course of 2 months from (September beginning to October end)

Initially we used this as a supplement to the Idiomatic Proficiency Test which contained a few more additions to it, to assess the participants level on usage of idioms in spontaneous situations. The method followed was: -

➤ Talking to the participants normally
➤ Suddenly giving them 3 options containing idiomatic expressions in correspondence to the conversation being held
➤ Assessing their response as per context

Later these meetings were changed from assessing to teaching perspective and the participants included majorly the school students. This method of giving examples of usage of idioms in conversations became one of the strongest and best methods we used to help the participants learn idioms and phrases.

We interviewed 25 students and each interview lasted between 15 and 20 minutes. At the beginning of each interview there was a brief conversation about their language of preference (English or Hindi) to help to create a relaxed atmosphere.

This was used primarily as supplementary / supportive evidence for the interpretation of the results of the Final IPT after training the participants in the three different methods.

## *Data Analysis*

SIMILAR AGE BUT DIFFERENT PROFESSIONS
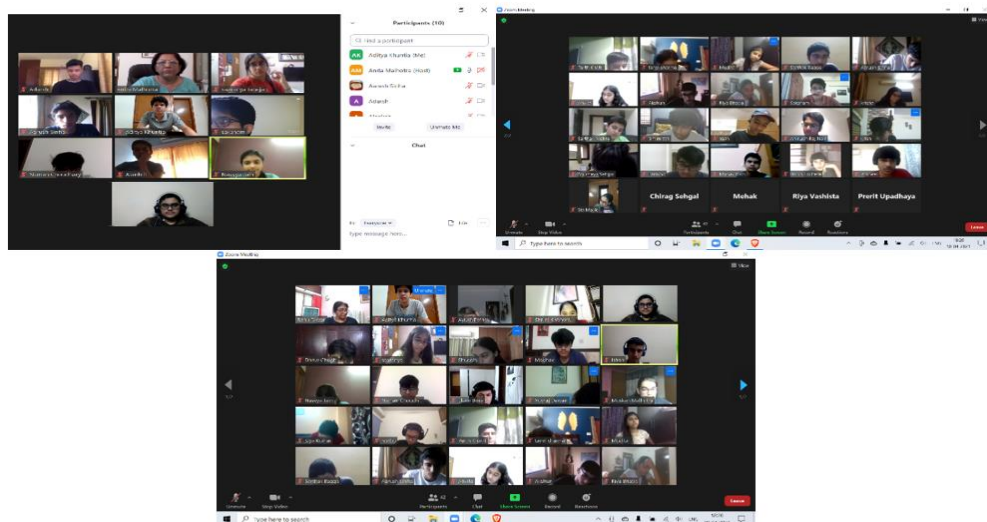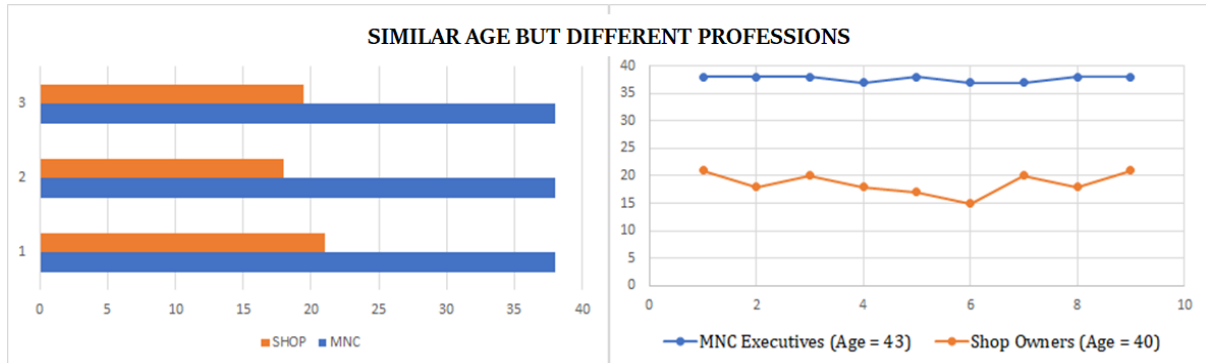
This graph shows the comparison in scores between people in different professions with the same age For this, we compared 2 shop owners' scores (Age 40) average with 3 MNC executive's scores (Age 43) averages and we saw that the MNC executives performed much better due to their exposure in these kinds of conversations


NATIVE VS NON NATIVE SPEAKERS
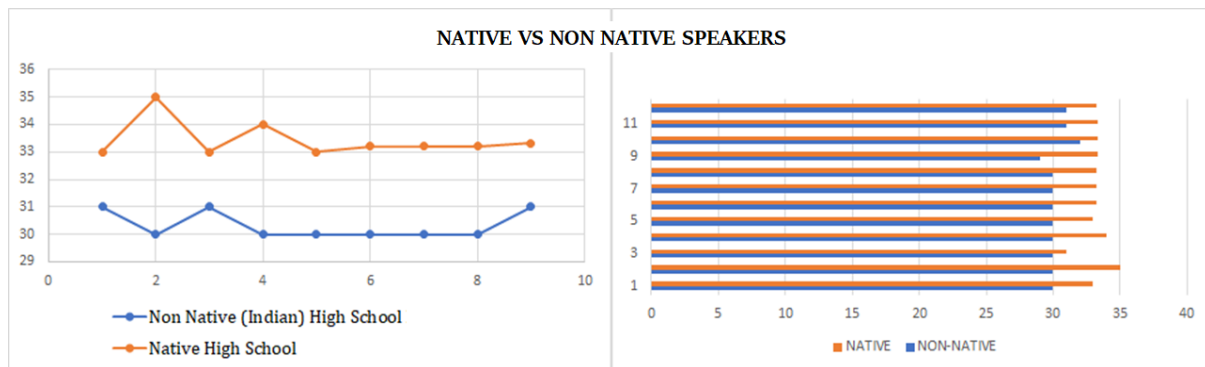
This graph shows the comparison between the test scores of native and non-native speakers. For this comparison, we took sample groups of American and Indian high school students and compared their scores and in every case we saw that the American students performed better than the Indian students with a margin of about 4-5 marks

SCORE AT DIFFERENT AGE
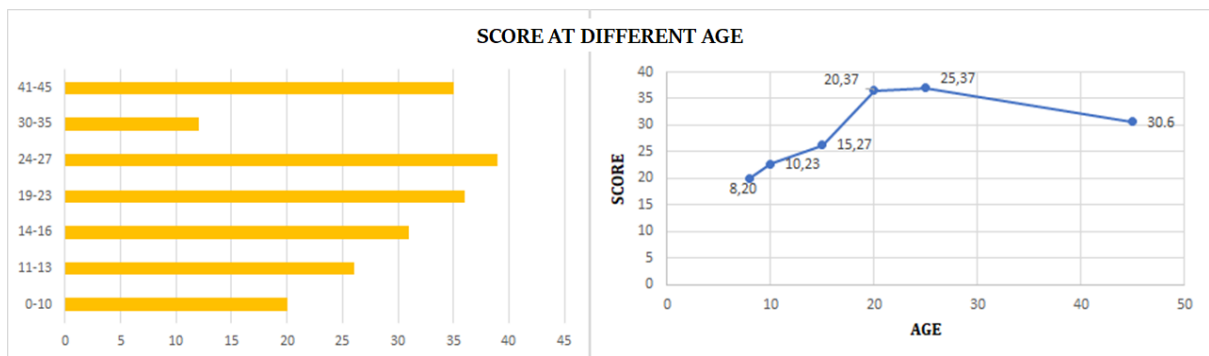
This graph shows the comparison between scores of different survey participants categorized based on age. We can see a progressive incline until the age group of 24-27 (Post graduate students). However, after that, the score suddenly drops as most people interviewed in the age range of 30-35 were either shop owners (10th level education) or people from rural areas (5th level education) which led to a much lower score. However, the score increases again in the 41-45 age range as that consisted mainly of MNC executives who got high scores but were still lower than the English major post-graduate students

COMPARISON BY LEVEL OF EXPERIENCE
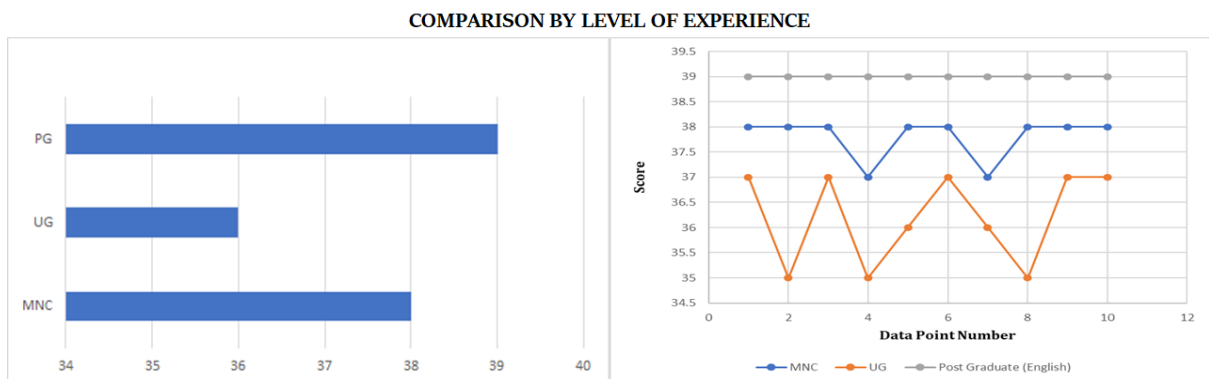
This graph shows the comparison between scores of different survey participants based on level of experience. For this graph, we took the scores of 3 types of individuals, MNC executives, post-graduate students (English major) and graduate level students. Through this data, we saw a large increase in scores from undergraduate students to postgraduate students. However, it decreased from postgraduate students to MNC executives since English major postgraduate students had a higher knowledge base of English.

COMPARISON ON THE BASIS OF EDUCATION LEVEL
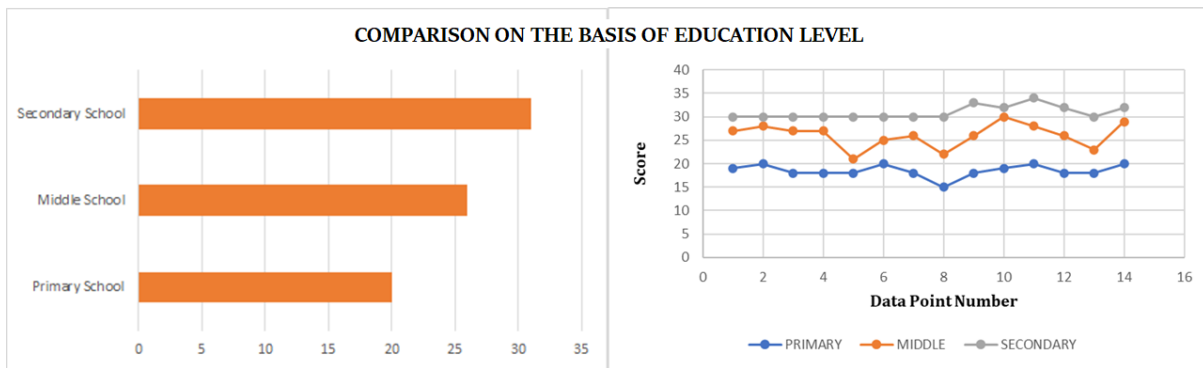
This graph shows the comparison between scores of individuals on the basis of education level and schooling. For this, we took the average scores of students from primary, middle, and secondary school and compared the results. We saw that there was a progressive increase in test scores with increase in education levels.

Therefore, we concluded that there were clear distinctions between:
➢ Native and non-native speakers
➢ Different education/schooling levels
➢ Different ages
➢ Different experiences
➢ Different professions

This led us to create different methods by which we could increase learning capacities of sample groups and helped us better understand the specifics of the problem and how we could find the solution.

After testing initial vocabulary using the IPT (Idiomatic Proficiency Test), participants were then divided into 3 groups with participants from every skill set in the groups. These groups were then given better understanding of idioms and their meanings using 3 different methods respectively:

➢ Showcasing just the Phrases and Idioms with their meanings
➢ Showcasing Phrases and Idioms in a context-based sentence.
➢ Showcasing the Phrases and Idioms in a model conversation

In the following weeks, the efficacy of these methods was tested using various methods such as tests, zoom calls etc. A Zoom call was set up bi-weekly with each of the groups and the following procedure was followed:

➢ Participants were firstly given training based on which sample set they were in.
➢ Then they were asked to give simple quizzes and tests
➢ These were then used to track their progress through one-month study.

Finally, after 8 bi-weekly meetings were held, participants were given another test to track their progress and conclusions were accordingly.
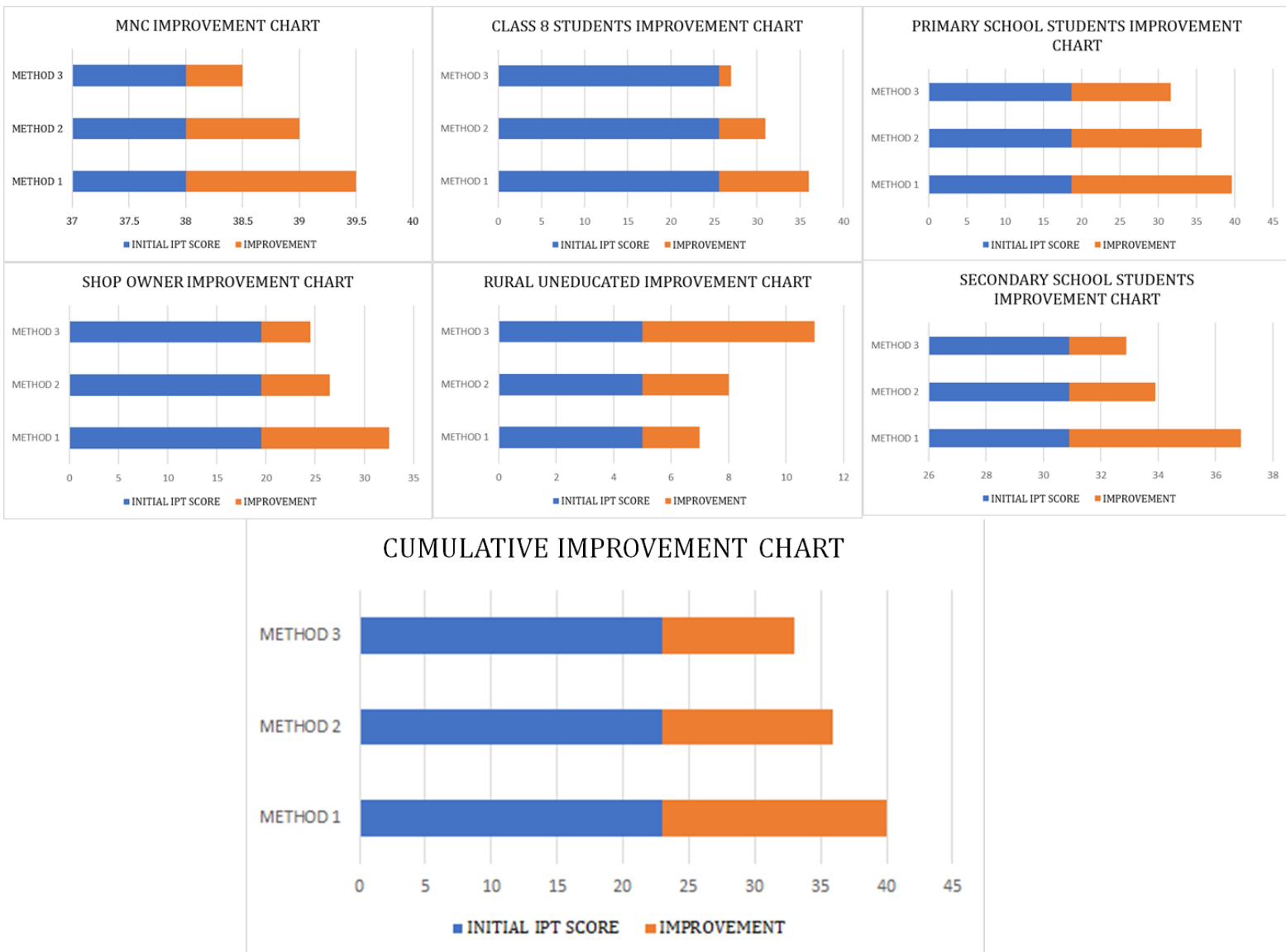
Final Results

Below are the charts showing improvement in usage of idiomatic sentences and phrases for expressing themselves of each category respectively, wherein

METHOD 1: Showing Idioms in a model conversation and thus figuring out context.
METHOD 2: Showing Idioms with a contextual usage in a sentence.
METHOD 3: Showing Idioms only with their meaning.

Thus, we can see that the usage of model conversations was the best among the three methods especially for growing children and primary students for whom model conversation gave them an all-round understanding of usage of global phrases and idioms.

For the Uneducated, the contextual examples became too difficult to understand and instead phrases in correspondence to their meaning were the best option.

For the MNC Executives, the method of showing idiom with its usage in a sample sentence also proved to be very effective as they can grasp things very quickly.

Thus, in our app we have included all three methods, with the main focus on usage of idioms in a model conversation of different scenarios. We have also included a Phrase Gallery which includes many idioms and phrases, with its respective meaning, pronunciation and used in a sentence to provide its contextual understanding.

THE PROTOTYPE

The prototype is available at https://eytoy2.pythonanywhere.com
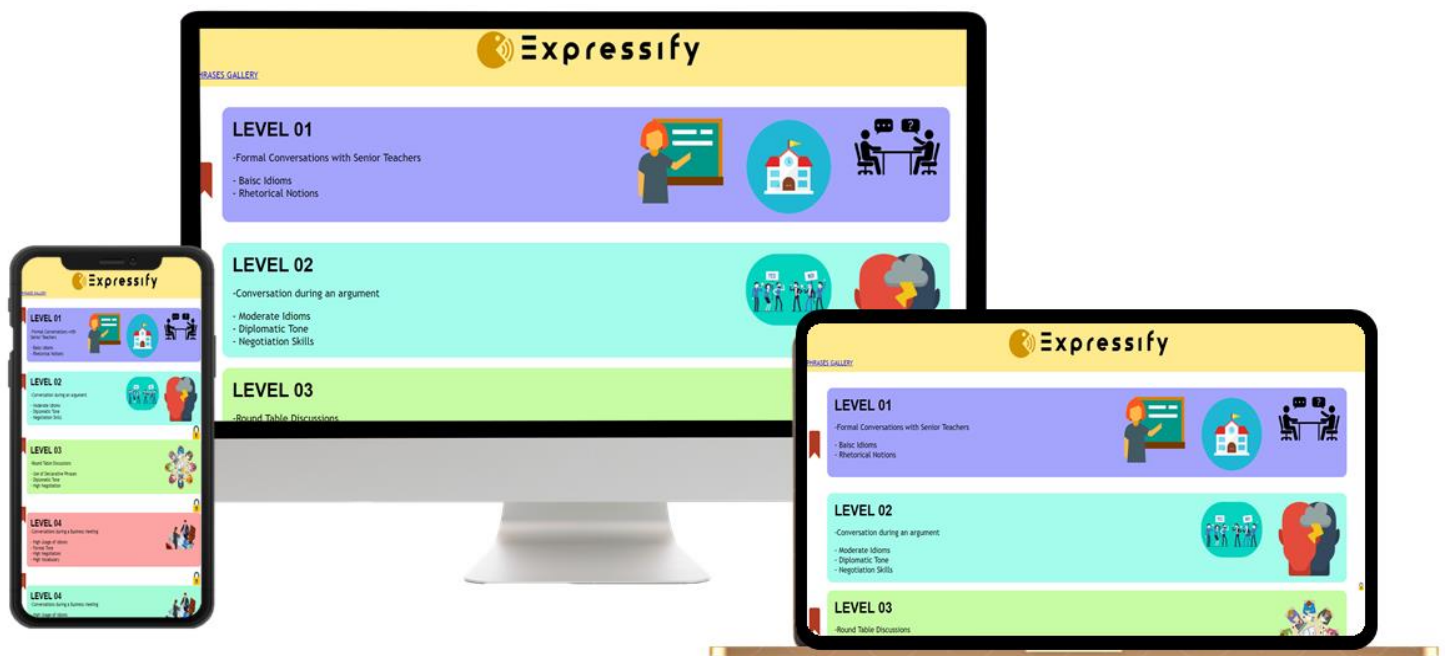
Expressify Model Conversations Homepage

The figure given above is an image prototype of our homepage. On the homepage, our application is divided into many levels. Each level comes with an increasing difficulty, higher vocabulary, global phrases and quiz formats with model conversations researched from both formal and informal scenarios while also considering real life scenarios which we too, have experienced.

The user at each challenge in a level is provided with an objective for their conversation. The goal is to reach that goal with maintaining the highest possible positivity scores.

Each of these objectives have been found after extensive research and surveys with data collected from previously held sample groups for comparison.

Expressify Model Conversations Quiz Format with Positivity Meter

Each challenge given on the homepage consists of a conversation with a bot where you are provided 3 choices and there are auto generated bot responses in accordance with the choices. Each choice made by the user affects the direction in which the conversation moves forward.

The Positivity Meter also judges the responses, their tone with respect to the bots auto generated sentence and finds out about the mood (i.e., How positive the conversation is) of the conversation. This is based upon a neural network *(Bi-directional LSTM (Long Short-Term Memory) Model)* in the backend and the users are given a bar-like output for easy understanding.

At the end of the conversation the user is provided with a sample flowchart which shows them the direction in which the conversation could have gone had they picked another option. This is to make them understand how to effectively communicate and how each word, phrase, tone of a response affects the conversation.

Expressify Phrase Gallery Page

This is an image prototype of our phrase gallery. The phrase gallery <u>includes commonly used idioms in conversation used by Native Speakers, which are not quite famous among the Non-Native Speakers</u>. These have also been used throughout the model conversations quiz given in the application. The gallery provides further knowledge regarding the meaning of the phrases and their usage in sentences and also provides the user with audio to help improve their skills in pronunciation of the words and phrases.

Furthermore, the <u>audio is given in both the local, accented English</u> which is due to the country of origin (Such as Latin *"Mea Culpa"*, French *"Au Courant"*, American English *"Cut the Mustard"* etc.) and non-accented English where it is pronounced without accents. This can help users better understand various different local accents and ways of speaking English to further their conversational skills.

Throughout the challenges, users are required to select these phrases in a random situation provided and using context modelling from the neural network in the backend, they are marked upon their understanding and usage.
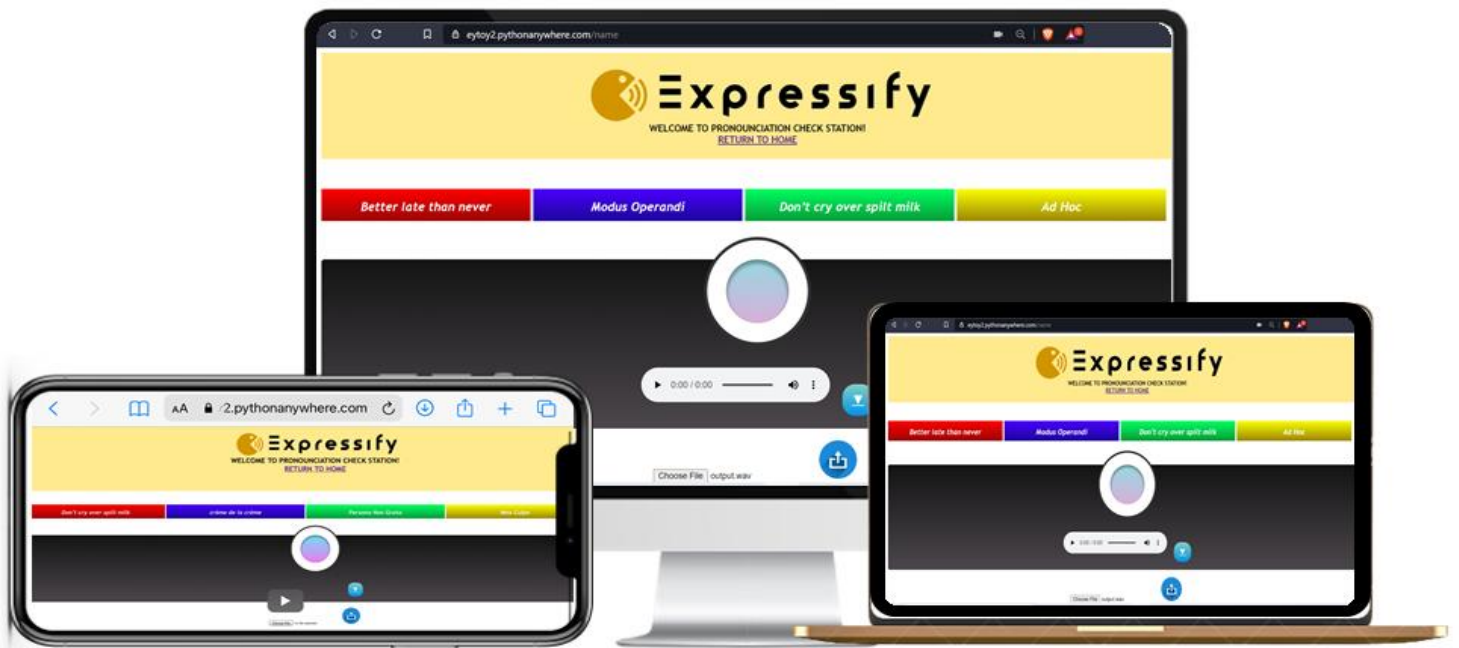
Expressify Pronunciation Check Station

This is the image prototype of our pronunciation check station. Here the user is provided with 4 phrases and a record button below. The user must speak the 4 phases in order and upload it below. Then the response audio is converted to text using Speech recognition and we compare the Question string and Response through the Levenshtein Distance which tells how similar are two strings and below is the result
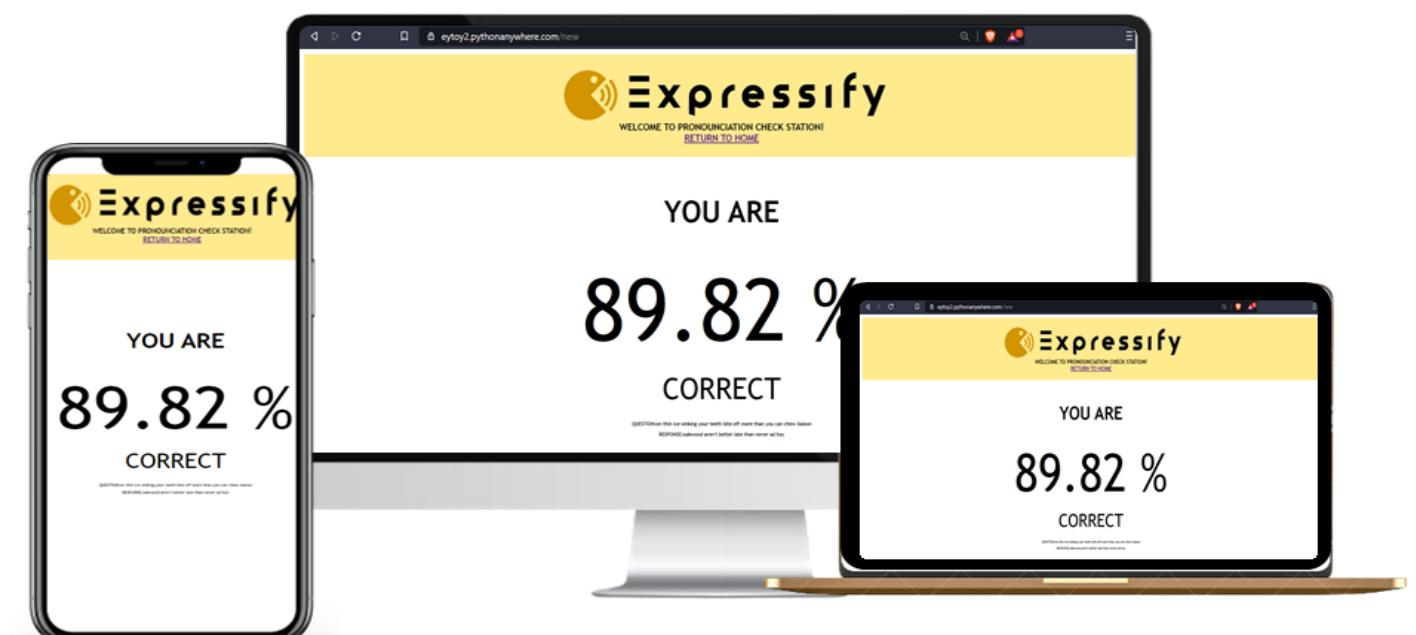
Expressify Pronunciation Check Result Page

ACKNOWLEDGEMENTS

We Would like to express our gratitude to

Amity International School, Saket: Our School for providing us with the opportunity to present our project in this competition.

PythonAnywhere: An online integrated development environment and web hosting service based on the Python programming language which allowed us to host our prototype website on the web for free.

Participants: of our Proof-of-Concept Study who took part in various surveys and calls with us and helped in testing of our prototype

---

******