پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

طراحی و پیاده سازی یک پلتفرم تحت وب برای جمع‌آوری دادگان صوتی

نگارنده

امیرمسعود آقایی

استاد راهنما: دکتر باقر باباعلی

پایان‌نامه برای دریافت درجه کارشناسی
در رشته علوم کامپیوتر

۸ اسفند ۱۴۰۱

چکیده

یکی از نیازمندي‌های اساسی برای توسعه یک سیستم پردازش و بازشناشی گفتار، در اختیار داشتن دادگان گفتاری با حجم و کیفیت مناسب است. دادگان گفتاری می‌بایست باید از منظر سن، جنسیت، میزان تحصیلات و لهجه گویندگان از تنوع کافی برخوردار باشد. در این پروژه هدف طراحی و پیاده سازی یک پلتفرم تحت وب برای جمع آوری دادگان صوتی بصورت انبوه از گویندگان مختلف است. این پلفترم این قابلیت را خواهد داشت که بتوان آن را برای سناریوهای مختلف جمع‌آوری گفتار با کاربردهای مختلف پیکربندی کرد و از رابط گرافیگی کاربرپسندی برخوردار خواهد بود.

سپاسگزاری

سپاسگزاری

از جناب آقای دکتر باقر باباعلی به‌عنوان استاد راهنما که همواره بنده را مورد لطف خود قرار داده‌اند، کمال تشکر را دارم؛ چرا که بدون راهنمایی‌های ایشان تامین این پایان‌نامه بسیار مشکل می‌نمود.

پیشگفتار

به خوبی شناخته شده است که سیستم‌های تشخیص خودکار گفتار مبتنی بر روش‌های آماری به مقادیر زیادی از داده‌های گفتاری رونویسی شده و حاشیه‌نویسی شده برای دستیابی به نرخ‌های دقت قابل قبول نیاز دارند. به‌خصوص هنگام پرداختن به زبان‌هایی با منابع کمتر یا حتی هر زبان دیگری که از نظر ارتباط اقتصادی بازار در بین پنج زبان بزرگ نیست (انگلیسی، اسپانیایی، فرانسوی، آلمانی و ایتالیایی) به‌دست آوردن تعداد زیادی داده گفتاری دشوار است. دلیل اصلی این امر این است که این مجموعه‌ها گران هستند و ثابت شده است که جذب سخنرانان بسیار پرهزینه[۱] و مدیریت آن دشوار است. علاوه بر این، برخی از پایگاه‌های داده گفتار به دلیل شرایط بد ضبط، ناهماهنگی نرخ نمونه، رونویسی اشتباه، ناسازگاری یا عدم وجود و دلایل دیگر کیفیت ندارند[۲] .

این مقاله راه حلی را برای مقابله با این موضوع با استفاده از رویکرد جمع‌سپاری [۱] توصیف می‌کند. جمع سپاری اصطلاحی است که برای توصیف استفاده از تعداد زیادی از مردم برای دستیابی به یک هدف خاص به روشی مشترک از طریق اینترنت استفاده می‌شود. بسیاری از طرح‌های جمع‌سپاری به دلیل در دسترس بودن فناوری‌های Web 2.0 امکان‌پذیر شده‌اند، که امکان انجام پروژه‌های همکاری گسترده را فراهم می‌کند. جمع سپاری را می توان به عنوان یک فرآیند توزیع شده برای حل مشکلات در نظر گرفت. به طور معمول فرآیند به شرح زیر است: یک واحد تجاری مشکلی دارد و باید آن را به روشی مقرون به صرفه حل کند. نهاد مشکل را در وب منتشر می‌کند و معمولاً ابزارهایی را برای حل آن ارائه می‌دهد. کاربران (جمعیت) به تماس پاسخ می‌دهند و راه حل‌هایی برای مشکل پیشنهاد می‌کنند. نهاد ناشر راه حل برنده را انتخاب می‌کند و بر این اساس به کاربر/کاربران پاداش می‌دهد. جوایز از مشوق‌های پولی گرفته تا فقط به رسمیت شناختن عمومی متفاوت است. نهاد ناشر صاحب راه حل برنده نهایی خواهد بود. راه‌حل‌های متعددی در سراسر وب یافت می‌شود تا رونویسی گفتار، طبقه‌بندی آهنگ، طبقه‌بندی کهکشان از بررسی آسمان اسلون، یافتن ایده برای مسائل پیشنهادی، برچسب گذاری تصویر و ویدئو و حتی خلاصه‌سازی ای از کل دانش بشری را انجام دهیم.[۳]

---

[۱] Crowd-sourcing

فهرست مطالب

فصل ۱

مفاهیم مقدماتی

۱.۱ سیگنال‌های گفتار

سیگنال‌های گفتاری می‌توانند انواع مختلفی از اطلاعات را در اختیار ما قرار دهند. این گونه اطلاعات عبارتند از:

- تشخیص گفتار، که اطلاعاتی در مورد محتوای سیگنال‌های گفتار می‌دهد.
- تشخیص گوینده که حاوی اطلاعاتی درباره هویت گوینده است.
- تشخیص احساسات، که اطلاعاتی را در مورد وضعیت عاطفی گوینده ارائه می‌دهد.
- تشخیص سلامت، که اطلاعاتی در مورد وضعیت سلامت بیمار ارائه می‌دهد.
- تشخیص زبان، که اطلاعات زبان گفتاری را به دست می‌دهد.
- تشخیص لهجه، که اطلاعاتی در مورد لهجه گوینده تولید می‌کند.
- تشخیص سن که اطلاعات مربوط به سن گوینده را ارائه می‌دهد.
- تشخیص جنسیت، که حاوی اطلاعاتی در مورد جنسیت گوینده است.

۱.۱.۱ تشخیص گفتار

تشخیص گفتار توانایی یک ماشین یا کامپیوتر برای تشخیص محتوای کلمات و عبارات در یک زبان گفته شده و تبدیل آن‌ها به یک قالب قابل فهم برای ماشین است. تشخیص گفتار را می‌توان در بسیاری از برنامه‌ها استفاده کرد. چنین برنامه‌هایی در موارد زیر ظاهر می‌شوند: دیکته کردن رایانه‌ها به جای تایپ کردن، کمک به افراد معلول، خانه‌های هوشمند و بسیاری موارد دیگر.

۲.۱.۱ تشخیص و تایید هویت گوینده

تشخیص گوینده را می‌توان به عنوان فرآیند شناسایی گوینده ناشناخته بر اساس اطلاعات تعبیه شده در سیگنال گفتار او با استفاده از ماشین (رایانه) تعریف کرد. تشخیص گوینده به دو بخش تقسیم می‌شود: شناسایی گوینده و تأیید هویت گوینده (احراز هویت). فرآیند تعیین اینکه یک گفته معین با کدام یک از گوینده‌های ثبت شده در رایانه مطابقت دارد، بخش شناسایی گوینده نامیده می‌شود. این قسمت را می‌توان در اماکن عمومی و یا برای رسانه‌ها استفاده کرد. این موارد شامل، اما نه محدود به، نهادهای دولتی، تماس با ایستگاه‌های رادیویی، آژانس‌های بیمه، یا مکالمات مستند است.[۴][۵] بخش راستی‌آزمایی گوینده به‌عنوان روشی برای پذیرش یا عدم پذیرش هویت سخنران ادعاشده توصیف می‌شود. کاربردهای این بخش شامل استفاده از صدا به عنوان یک عامل کانونی برای تأیید هویت گوینده ادعا شده است. روابط تجاری با استفاده از شبکه تلفن، امکانات دسترسی به مجموعه داده ها، کنترل امنیتی برای محافظت از اطلاعات خصوصی، دسترسی از راه دور به رایانه و سیستم های‌مراقبت بهداشتی هوشمند از حوزه های کاربردی این شاخه است.[۶][۷]

۳.۱.۱ تشخیص احساسات

تشخیص احساسات توسط ماشین را می‌توان به عنوان وظیفه تشخیص احساسات ناشناخته بر اساس اطلاعات درج شده در سیگنال‌های گفتاری تعریف کرد. زمینه تشخیص احساسات به دو شاخه شناسایی احساسات و تأیید احساسات تقسیم می‌شود. در شاخه اول، احساس ناشناخته به عنوان احساسی شناخته می‌شود که مدل آن به بهترین وجه با سیگنال گفتار ورودی مطابقت دارد. در شاخه دوم، هدف تعیین این است که آیا یک عاطفه به یک عاطفه شناخته شده خاص تعلق دارد یا به برخی از احساسات ناشناخته دیگر. کاربردهای تشخیص احساسات به وضوح در [۸][۹] درک وضعیت عاطفی گوینده در مکالمات مرکز تماس تلفنی و ارائه بازخورد به اپراتور برای مشاهده اهداف، طبقه‌بندی پیام‌های پست صوتی بر اساس احساسات بیان شده توسط تماس‌گیرندگان و تشخیص افراد بی‌اعتماد ظاهر می‌شوند.

۴.۱.۱ تشخیص سلامت بیمار

تشخیص خودکار سلامت به عنوان استفاده از صدای بیمار برای ارائه اطلاعات در مورد وضعیت سلامت بیمار تعریف می‌شود. تشخیص خودکار سلامت می‌تواند در سیستم‌های مراقبت بهداشتی هوشمند استفاده شود.[۶][۷] این سیستم‌ها را می‌توان در بیمارستان‌هایی که شامل روش‌های طبقه‌بندی و ارزیابی سلامت رایانه‌ای می‌شود، استفاده کرد[۶]. این سیستم‌ها همچنین می‌توانند در ارزیابی صدای آسیب شناختی (صداهای ناهنجار عملکردی) استفاده شوند[۷] . صدای ناهماهنگ می‌تواند خشن یا بسیار نفس‌گیر باشد. علاوه بر این، از سیستم‌های تشخیص سلامت خودکار می‌توان

در تشخیص بیماری پارکینسون استفاده کرد. در موسسه فناوری ماساچوست[1] ،(MIT) تیمی به رهبری مکس لیتل آزمایش‌ها را برای تجزیه و تحلیل و ارزیابی ویژگی‌های صوتی بیمارانی که به بیماری پارکینسون شناسایی شده بودند، انجام دادند. آنها دریافتند که می‌توانند ابزاری برای تشخیص چنین بیماری در الگوهای گفتاری افراد ایجاد کنند.

۵.۱.۱ تشخیص زبان

تشخیص زبان مسئله تعیین زبان طبیعی است که محتوای گفتاری داده شده در آن قرار دارد. یکی از چالش‌های عظیم سیستم‌های تشخیص زبان، تمایز بین زبان‌های نزدیک به هم است. زبان‌های مشابه مانند صربی و کرواتی یا اندونزیایی و مالایی همپوشانی واژگانی و ساختاری قابل توجهی را نشان می‌دهند. از این رو، تشخیص تفاوت بین هر دو زبان برای سیستم‌های تشخیص زبان چالش برانگیز می‌شود. کاربردهای تشخیص خودکار زبان به وضوح در ترجمه زبان گفتاری [۱۰] ، تشخیص گفتار چند زبانه [۱۱] و بازیابی اسناد گفتاری [۱۲] ظاهر می‌شود.

۶.۱.۱ تشخیص لهجه

وظیفه تشخیص لهجه، تشخیص لهجه براساس منطقه محل زندگی گوینده، در یک زبان از پیش تعیین شده، تنها با توجه به سیگنال صوتی است. مشکل تشخیص لهجه به دلیل شباهت بیشتر بین لهجه‌های یک زبان، چالش برانگیزتر از تشخیص زبان در نظر گرفته شده است [۱۳]. تشخیص لهجه طیف وسیعی از کاربردها را در زندگی روزمره ما دارد. تشخیص لهجه به تشخیص خودکار گفتار (ASR) کمک می‌کند، زیرا گوینده‌هایی با لهجه‌های متنوع، برخی از کلمات را متفاوت تلفظ می‌کنند. تشخیص لهجه همچنین به ما این امکان را می‌دهد که منشاء و قومیت منطقه‌ای گوینده را نتیجه‌گیری کنیم و در نتیجه ویژگی‌های مورد استفاده در تشخیص گوینده را با منشاء منطقه‌ای تنظیم کنیم.

۷.۱.۱ تشخیص سن

تشخیص سن از طریق صدا، فرآیند تخمین سن گوینده(به عنوان مثال کودک، جوان، بزرگسال، سالمند و غیره) با استفاده از سیگنال‌های گفتاری او است. تشخیص سن خودکار را می‌توان در برنامه‌های امنیتی، برنامه‌های کاربردی محدودیت سنی و موارد دیگر استفاده کرد[۱۴].

۸.۱.۱ تشخیص جنسیت

تشخیص جنسیت خودکار فرآیند تشخیص مرد یا زن بودن گوینده است. به طور کلی، تشخیص جنسیت خودکار بدون تلاش زیاد منجر به دقت بسیار زیادی می شود، زیرا نتایج این نوع تشخیص

[1] Massachusetts Institute of Technology

دودویی است(چه مرد و چه زن) [۱۵][۱۶] تشخیص جنسیت خودکار به وضوح در کاربردهای مراکز تماس برخی از جوامع محافظه کار ظاهر می‌شود، جایی که سیستم‌های گفت‌وگوی خودکار با توانایی تشخیص جنسیت بر سیستم‌هایی که چنین توانایی ندارند ترجیح داده می‌شود.

۲.۱ جمع‌آوری دادگان صوتی برای تشخیص گفتار

از ماشین‌های خودکار گرفته تا تشخیص مراقبت‌های بهداشتی، برنامه‌های تشخیص گفتار/صدا پیشرفت‌هایی را برای بسیاری از صنایع به ارمغان می‌آورند. داده‌ها بخشی جدایی ناپذیر از توسعه و بهبود سیستم‌های تشخیص گفتار است زیرا سطح کلی عملکرد سیستم به کیفیت داده‌ها بستگی دارد. جمع‌آوری یا تولید چنین داده‌هایی می‌تواند دشوار باشد، به خصوص اگر از روش مناسب استفاده نشود.

۱.۲.۱ روش‌های برتر جمع آوری دادگان صوتی

مجموعه دادگان صوتی از پیش بسته‌بندی شده

مجموعه دادگان صوتی از پیش بسته‌بندی شده[۲] برای توسعه و بهبود مدل‌های تشخیص گفتار اولیه مناسب هستند. آن‌ها مجموعه داده‌های آماده‌ای هستند که به صورت آنلاین برای خرید از فروشندگان مختلف در دسترس هستند.

مزایا

- در مقایسه با جمع‌آوری دادگان صوتی داخلی، ارزان تر است.
- این مجموعه دادگان بزرگ هستند و می‌توان آن‌ها را به سرعت خریداری کرد.
- کیفیت نسبتاً بهتر از مجموعه دادگان صوتی عمومی است زیرا شرکت‌ها این مجموعه‌ها را به جای عموم مردم ضبط می‌کنند.

[۲] Prepackaged voice datasets

معایب

- چنین مجموعه داده‌هایی قبل از استفاده، نیاز به پیش پردازش قابل توجهی دارند که هزینه‌های پردازش را به بودجه اضافه می‌کند.
- آنها نمی‌توانند موارد استفاده خاص و منحصر به فرد از پروژه‌های تشخیص گفتار را پوشش دهند.
- این مجموعه داده‌ها قابل تنظیم/مقیاس‌پذیری نیستند و افزودن داده‌های اضافی دشوار است.
- مدل‌های تشخیص گفتار مدرن در حال پیچیده‌تر شدن هستند، به ویژه مدل‌هایی که از یادگیری عمیق استفاده می‌کنند.[۱۷] مجموعه دادگان صوتی از پیش بسته‌بندی شده نمی‌تواند چنین الزاماتی را برآورده کند.

مجموعه دادگان صوتی عمومی

مجموعه دادگان از پیش بسته‌بندی شده مشابه مجموعه دادگان صوتی عمومی[۳] هستند. تنها تفاوت آن‌ها این است که مجموعه دادگان عمومی معمولاً رایگان هستند و سطح کیفیت و ویژگی بسیار پایین تری را ارائه می‌دهند. هدف از ایجاد مجموعه دادگان عمومی، حمایت از نوآوری در صنعت تشخیص گفتار است.

جمع آوری دادگان صوتی از مشتریان

این روش[۴] دیگری برای جمع‌آوریدادگان صوتی است که معمولاً توسط برندها استفاده می‌شود. این برندها معمولاً راه حل‌های مبتنی بر تشخیص گفتار مانند دستگاه‌های خانه هوشمند یا دستیارهای مجازی ارائه می‌دهند.

مزایا

- دادگان صوتی جمع‌آوری شده از مشتریان ارزان‌تر است و به وفور در دسترس است. فقط هزینه اولیه جمع‌آوری وجود دارد، اما مابقی با استفاده مشتری از محصول جمع‌آوری می‌شود.
- داده‌های صوتی تازه در کمترین زمان ممکن در دسترس هستند.
- دادگان صوتی با جزییات دقیق مورد استفاده قرار می‌گیرند زیرا مستقیماً از مشتریان جمع‌آوری می‌شوند. این باعث می شود داده‌های صوتی بسیار دقیق باشد.

[۳] Public voice datasets
[۴] Customer voice data collection

معایب

- دادگان صوتی نوعی داده بیومتریک هستند، به همین دلیل است که جمع آوری دادگان صوتی از مشتری در چند سال گذشته بحث برانگیز شده است. به دلیل حریم خصوصی و تهدیدات امنیتی، مشتریان معمولاً تمایلی به اشتراک گذاری داده‌های صوتی خود ندارند.
- بسیاری از کشورها در حال حاضر محدودیت‌های قانونی را برای جمع آوری داده‌های مشتری اعمال می‌کنند که می‌تواند استفاده از این روش را دشوار کند.

جمع‌آوری دادگان صوتی داخلی

جمع‌آوری دادگان صوتی داخلی[۵] نیز می‌تواند راهی برای ایجاد مجموعه داده‌های با کیفیت بالا و منحصر به فرد باشد. این روش برای پروژه‌هایی مناسب است که به مجموعه داده‌های بزرگ در چندین زبان یا گویش نیاز ندارند.

مزایا

- این روش برای پروژه‌های تشخیص صدا مخفی مانند ارتش مناسب است.[۱۸]
- آنها کنترل بیشتری بر فرآیند جمع‌آوری دادگان صوتی می‌دهند، به این معنی که توسعه دهنده می‌تواند انتخاب کند که از کدام دستگاه‌ها استفاده کند و چگونه نویز پس زمینه ضبط را کنترل کند.

معایب

- این روش می‌تواند پرهزینه باشد زیرا شامل استخدام مشارکت کنندگان، خرید تجهیزات ضبط، راه اندازی استودیو (در صورت لزوم) و غیره است.
- این روش می‌تواند برای جمع‌آوری مجموعه داده‌های متنوع دشوار باشد..
- از آنجایی که داده‌های صوتی در زمان واقعی جمع آوری می‌شوند، انجام آن در خانه می‌تواند تاخیرهای قابل توجهی را به جدول زمانی پروژه شما اضافه کند.

[۵] In-house voice data collection

روش جمع‌سپاری جمع‌آوری دادگان صوتی

اگر شرکتی مایل به تحمل دردسرهای مدیریت جمع‌آوری داده‌ها که خود یک پروژه است، نباشد، می‌تواند جمع سپاری[۶] را برون‌سپاری کند. اگر داده‌ها به چندین زبان و گویش مورد نیاز است، شرکت می‌تواند با یک ارائه‌دهنده خدمات جمع‌سپاری شخص ثالث متخصص در جمع‌آوری/حاشیه‌نویسی داده کار کند.

مزایا

- می‌توانید مجموعه داده‌های صوتی را با تعیین نیازهای خود به ارائه‌دهنده خدمات، سفارشی سازی و مقیاس کنید.
- از آنجایی که جمع سپاری از طریق یک برنامه آنلاین انجام می‌شود و مشارکت کنندگان از تجهیزات ضبط خود استفاده می‌کنند، می‌تواند ارزان‌تر از جمع‌آوری داده‌های صوتی داخلی باشد.
- از آنجایی که جمع‌سپاری طیف گسترده‌ای از مشارکت کنندگان را ارائه می‌دهد که در سراسر جهان پراکنده شده‌اند، داده‌های صوتی را می‌توان به چندین زبان و گویش جمع‌آوری کرد.
- ارائه دهندگان خدمات شخص ثالث نیز خدمات اضافی مانند پیش پردازش پس پردازش دادگان صوتی را ارائه می‌دهند.
- از آنجایی که داده‌های صوتی به عنوان داده‌های بیومتریک در نظر گرفته می‌شوند، داشتن مالکیت قانونی آن مهم است. فروشندگان شخص ثالث همچنین حقوق دادگان صوتی را انتقال می‌دهند تا به شما در جلوگیری از مشکلات قانونی آینده کمک کنند.

معایب

- از آنجایی که دادگان از راه دور از طریق تلفن‌های هوشمند یا سایر تجهیزات ضبط شخصی مشارکت کننده جمع آوری می‌شود، گزینه‌های کمتری از نظر انتخاب تجهیزات وجود دارد.
- این روش می‌تواند برای جمع‌آوری مجموعه داده‌های متنوع دشوار باشد..
- گاهی اوقات، دادگان صوتی با نویز خاصی در پس زمینه برای آموزش تشخیص صدا به منظور جلوگیری از نویز پس زمینه مورد نیاز است. برای رسیدن به این هدف، انواع دیگر نویز پس زمینه باید حذف شوند. با این حال، پاک کردن صدای پس زمینه می‌تواند چالش برانگیز باشد زیرا همه مشارکت کنندگان به استودیوهای ضبط یا اتاق‌های عایق صدا دسترسی ندارند. بنابراین، قبل از کار با یک فروشنده، مطمئن شوید که چنین مشخصاتی را ارائه می‌دهد.

[۶] Crowdsourcing voice data collection

ما در این پروژه، سعی کردیم تا با استفاده از روش جمع‌سپاری یک پلتفرم جمع‌آوریدادگان صوتی تحت وب با رابط کاربری و گرافیکی مناسب ارائه دهیم. در فصل‌های آتی درباره چگونگی عملکرد این پلتفرم و همچین نحوه عملکرد آن و ابزارهای استفاده شده صحبت می‌کنیم.

فصل ۲

طراحی پلتفرم

پلتفرم طراجی شده یک وب اپلیکیشن برای جمع‌آوری دادگان صوتی از کاربر است. در این پلتفرم از ابزارهای طراحی و ساخت وبسایت در بخش فرانت و بک کمک گرفته شده است و سعی شده تا سیستم رابط کاربری مناسب داشته باشد تا کاربر در هنگام استفاده از آن احساس راحتی کند. در این فصل چگونگی عملکرد سیستم را توصیف می‌کنیم، اهداف جمع‌آوری دادگان صوتی در این پروژه را بیان می‌کنیم و در نهایت نحوه استفاده از این پلتفرم را به شما آموزش می‌دهیم.

۱.۲ توصیف عملکرد سیستم

در این پروژه، یک فریم ورک طراحی شده‌است که هدف آن جمع‌آوری دادگان صوتی از کاربر است، این دادگان در سرور ذخیره می‌شوند و سپس از این دادگان برای پیاده‌سازی یک مدل تشخیص گفتار بوسیله یادگیری ماشین و هوش مصنوعی استفاده می‌شود. هر کاربر در این پلتفرم اطلاعات مختلفی را قبل از ارسال گفتار ضبط شده خود در اختیار ما قرار می‌دهد. برخی از این اطلاعات که برای ما حیاتی هستند عبارتند از:

- جنسیت
- سن
- گویش
- میزان نویز فضای اطراف کاربر هنگام ضبط
- نام و نام خانوادگی

سیستم برای هر کاربر بر اساس اهداف مورد نظر سناریوی مربوطه مجموعه آزمون‌هایی را طراجی می‌کند. ممکن در بعضی از این آزمون‌ها برخی از اطلاعات فوق مانند جنسیت یا سن مهم نباشد و در برخی دیگر گویش فرد برای ما اهمیتی نداشته باشد. کاربر این مجموعه آزمون‌ها را یک به یک انجام می‌دهد، صدای خود را ضبط می‌کند و سپس با دکمه ارسال، صدای ضبط شده خود را برای سرور می‌فرستد. در هر آزمون سیستم یک فایل HTML را از سرور می‌گیرد و به کاربر نشان می‌دهد. سپس پس از ضبط صدا توسط کاربر، سرور دادگان را از کاربر دریافت کرده و آن را در پایگاه داده ذخیره می‌کند. همچنین سیستم شامل یک قسمت پنل برای ادمین سیستم می‌باشد. ادمین سایت توانایی مشاهده و حتی ویرایش اطلاعات کاربران را دارد و همچنین می‌تواند تا آزمون‌های جدیدی را برای کاربر تعریف کند. ادمین می‌تواند عملکرد کاربر مانند آزمون‌های انجام شده، باقی مانده و غیره را نیز مشاهده کند. همچنین اطلاعاتی راجع به هر آزمون بخصوص، نظیر اینکه این آزمون توسط چه تعداد کاربری انجام شده است بدست می‌آورد.

۲.۲ اهداف جمع‌آوری دادگان صوتی

چارچوب پیشنهادی راه‌حلی برای مجموعه‌ای از صوت‌های ضبط شده در مقیاس بزرگ برای آموزش مدل‌های یادگیری ماشین و هوش مصنوعی برای تشخیص گفتار، تشخیص احساسات گوینده، تشخیص هویت و غیره و تنظیم دقیق آن‌ها با مجموعه داده‌های شخصی‌سازی شده کوچک‌تر ارائه می‌کند. به طور خاص، یک پلتفرم ساخته شده‌است که یک روش استاندارد هدایت شده برای ضبط گفتار را اجرا می‌کند. کاربران می‌توانند دستورالعمل‌ها را دنبال کنند تا مجموعه داده شخصی‌شده گفتار را از طریق مرورگر ضبط کنند. این برنامه روشی از حاشیه‌نویسی دادگان در زمان ضبط و همچنین یک مرحله خودکار پس از پردازش تمام ضبط‌ها را ارائه می‌دهد. مدل‌های از پیش آموزش دیده برای یادگیری انتقال و تنظیم دقیق برای نیازهای SR استفاده می‌شود. مدل‌ها بر روی مجموعه داده‌ها در مقیاس بزرگ برای تعمیم بهتر آموزش داده می‌شوند و سپس با داده‌های خاص سخنران سازگار می‌شوند. بر اساس بررسی ادبیات ارائه شده، انتظار می‌رود که این امر عملکرد را در مجموعه تست، بدون تطبیق بیش از حد با داده‌ها، بهبود بخشد.

پس از رویکرد جمع‌سپاری، پروژه باید منافع متقابلی را هم برای تیم تحقیقاتی و هم برای جمعیت فراهم کند. استفاده از این پلتفرم امکان جمع‌آوری دادگان حاشیه‌نویسی شده در مقیاس بزرگ را فراهم می‌کند. این می‌تواند نتایج تحقیقات را در آینده، در مورد انواع مدل‌های تشخیص گفتار، توسعه دهد. هر مجموعه داده جدیدی که ارسال می‌شود، حاوی ضبط‌هایی از یک گوینده خاص است، برای تنظیم دقیق، از مدل مستقل از گوینده از پیش آموزش‌دیده استفاده می‌شود که منجر به یک مدل تشخیص گفتار سازگار با گوینده می‌شود. اکثر مجموعه دادگان موجود در شرایط ایده آل ضبط استودیویی ایجاد می‌شوند. انتظار می‌رود کاربرانی که ضبط‌های خود را با برنامه وب ارائه شده ارسال می‌کنند، از تجهیزات داخلی رایانه یا تلفن خود در شرایط غیر ایده‌آل ضبط خانگی

استفاده کنند. این می تواند منجر به یک مجموعه داده چند منبعی شود که برای برنامه‌های کاربردی تشخیص گفتار در دنیای واقعی که باید با تجهیزات غیر حرفه ای نگهداری شوند، موثرتر عمل می‌کنند. علاوه بر این، معماری‌های عمیق‌تر با ظرفیت یادگیری بیشتر می‌توانند بدون خطر برازش بیش از حد اجرا شوند. در نهایت، از آنجایی که برنامه وب از مشارکت چند زبانه پشتیبانی می‌کند، امکان انتقال دانش بین زبان‌های مختلف قابل بررسی است.

این چارچوب همچنین با هدف ارائه منافع به جمعیت است. اولاً، این یک پروژه دانشگاهی غیرتجاری با هدف واحد ارائه ابزار، نتایج تحقیقات و مجموعه داده عمومی برای جامعه است. این به عنوان یک انگیزه درونی برای افرادی که مایل به مشارکت در تحقیقات علمی هستند عمل می‌کند. این وب سایت بخشی را برای قدردانی از همه مشارکت کنندگان فراهم می‌کنند تا آنها بتوانند این موضوع را در رزومه و کارنامه خود بگنجانند. همچنین طراحی شده است تا یک فعالیت سرگرم کننده باشد که به تعهد زیادی نیاز ندارد. در نهایت، مدل‌های تشخیص گفتار شخصی‌سازی شده ایجاد شده، می‌توانند برای چندین برنامه خلاقانه استفاده شوند.

۳.۲ آموزش استفاده از وب اپلیکیشن

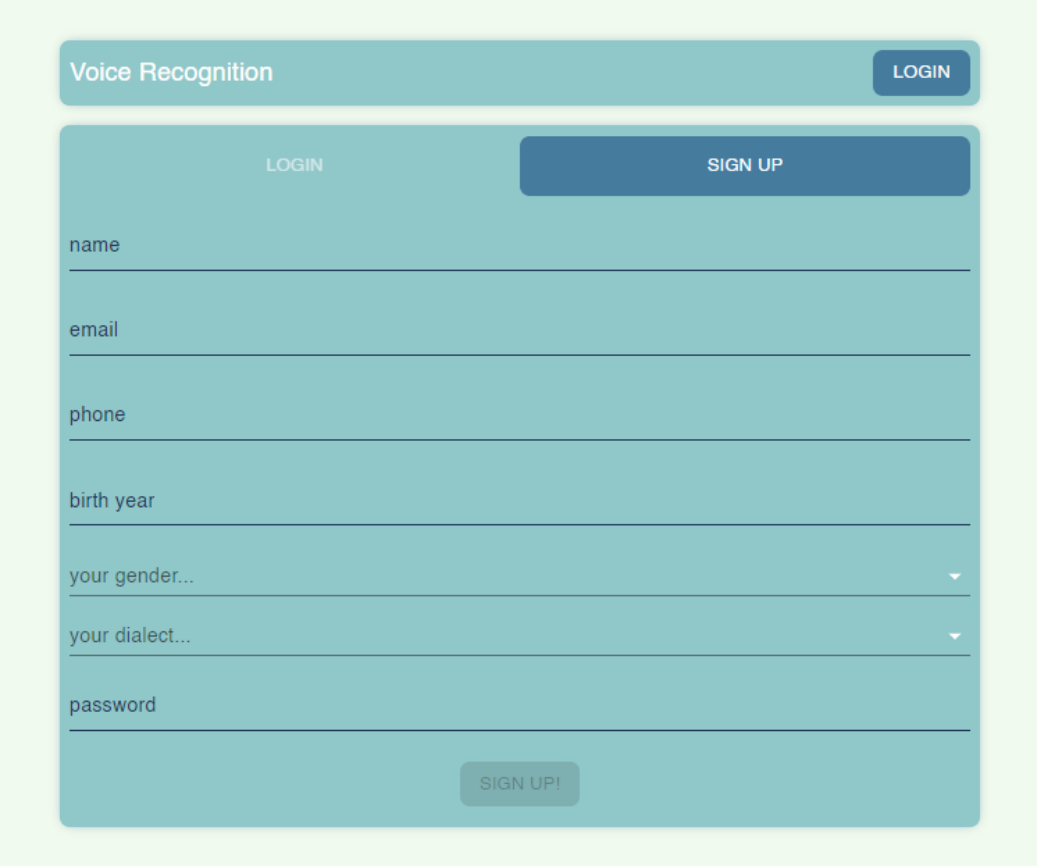
در این بخش یک آموزش استفاده از این وب اپلیکیشن ساده ارائه می‌شود تا نجوه کار پلتفرم به صورت بصری نیز نمایش داده شود.

۱.۳.۲ ثبت نام و ورود

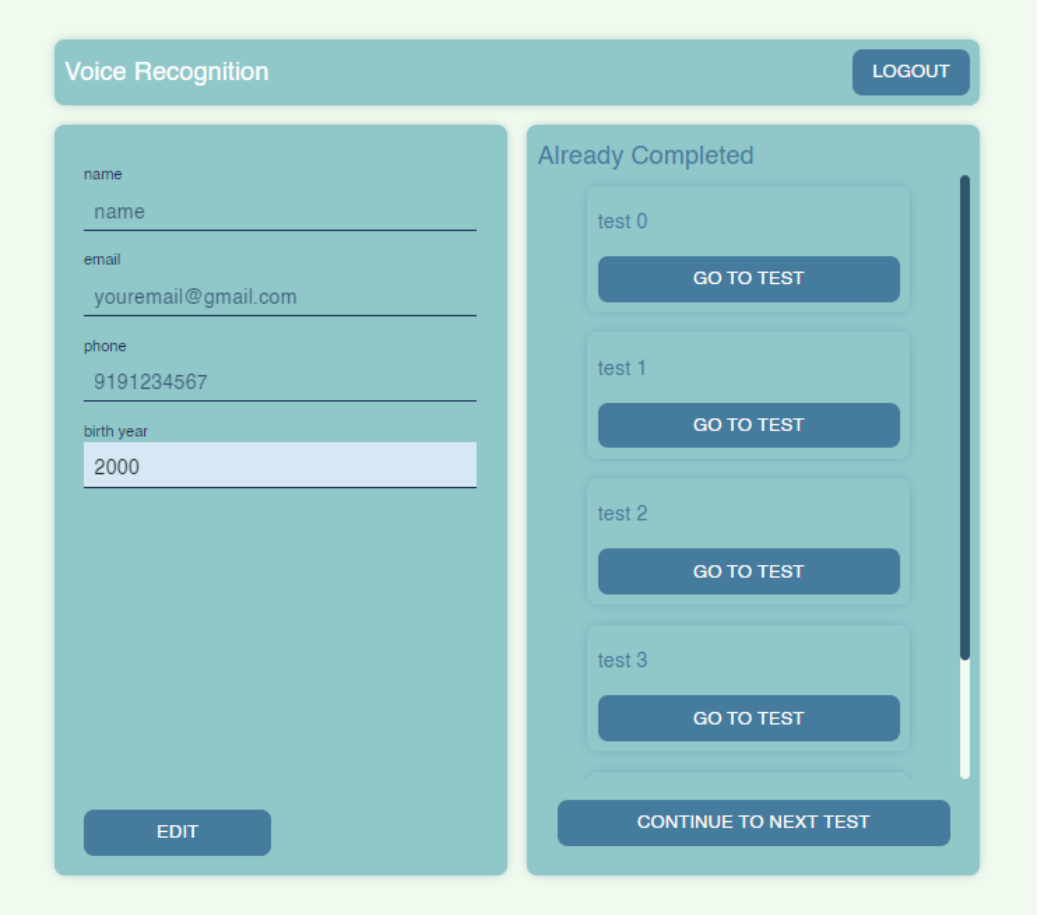
در این صفحه شما می‌بایست با وارد کردن اطلاعات خود یک حساب کاربری برای خود بسازید. اطلاعات شما می‌بایست صادقانه و دقیق باشد چرا که در ارزیابی نهایی مدل تشخیص گفتار تاثیر مستقیم دارد.
اگر قبلا ثبت‌نام کرده‌اید می‌توانید با انتخاب کردن گزینه login و وارد کردن نام کاربری و رمز عبور خود وارد حساب کاربری خود شوید.

۲.۳.۲ مشاهده آزمون‌ها

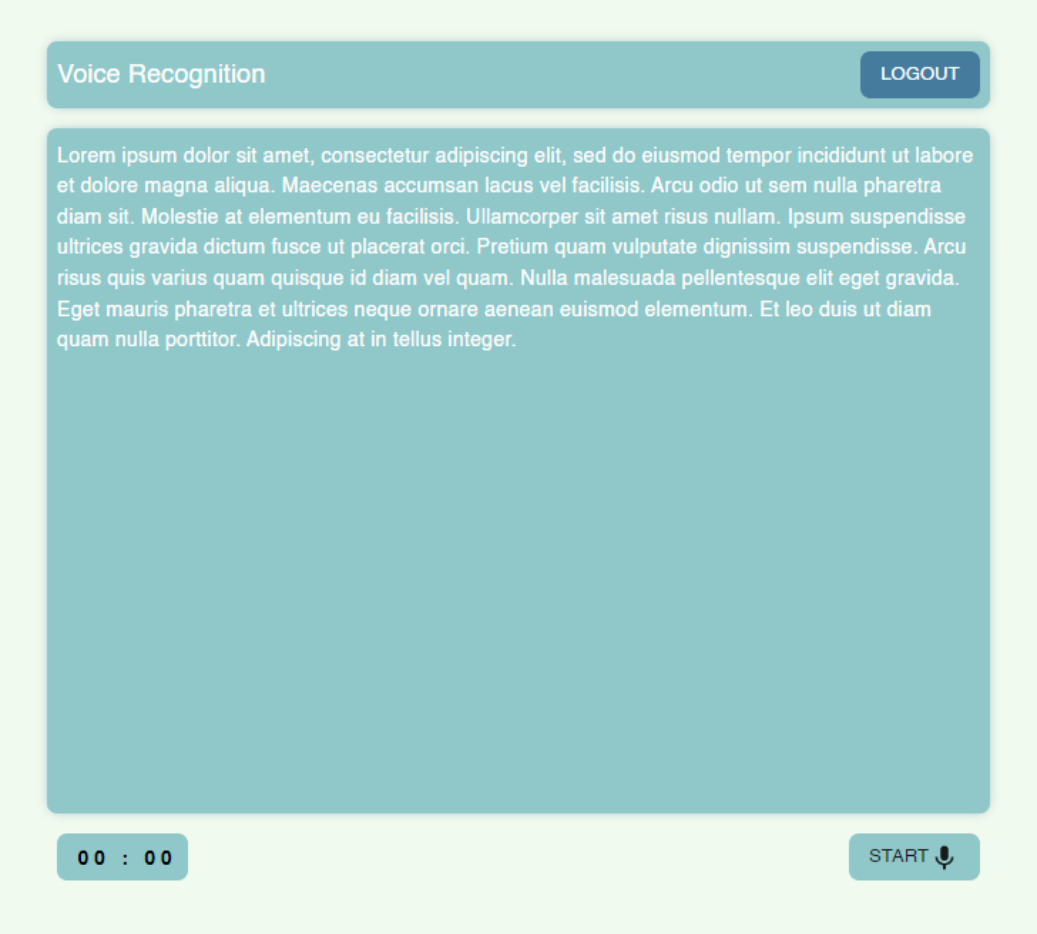
پس از ورود به حساب کاربری خود شما می‌توانید آزمون‌های تعریف شده را در نیمه سمت راست تصویر مشاهده کنید. با کلیک بر روی گزینه GO TO TEST می‌توانید وارد صفحه تست مربوطه شوید.
همچنین با انتخاب گزینه EDIT می‌توانید برخی از اطلاعات کاربری خود را ویرایش کنید.



شکل ۱.۲: صفحه ثبت نام و ورود کاربر



شکل ۲.۲: صفحه حساب کاربری



شکل ۳.۲: انجام مجموعه وظایف تعریف شده

۳.۳.۲ انجام وظایف تعریف شده برای کاربر

پس از انتخاب گزینه GO TO TEST وارد صفحه‌ای می‌شوید که باید مطابق با دستورالعمل مشاهده‌شده در روی صفحه عمل کرده و تست را انجام دهید.
توجه کنید که کاربر موظف است برای انجام دادن هر آزمون، آزمون‌های قبلی را تکمیل کرده باشد.

۴.۳.۲ ضبط صدا

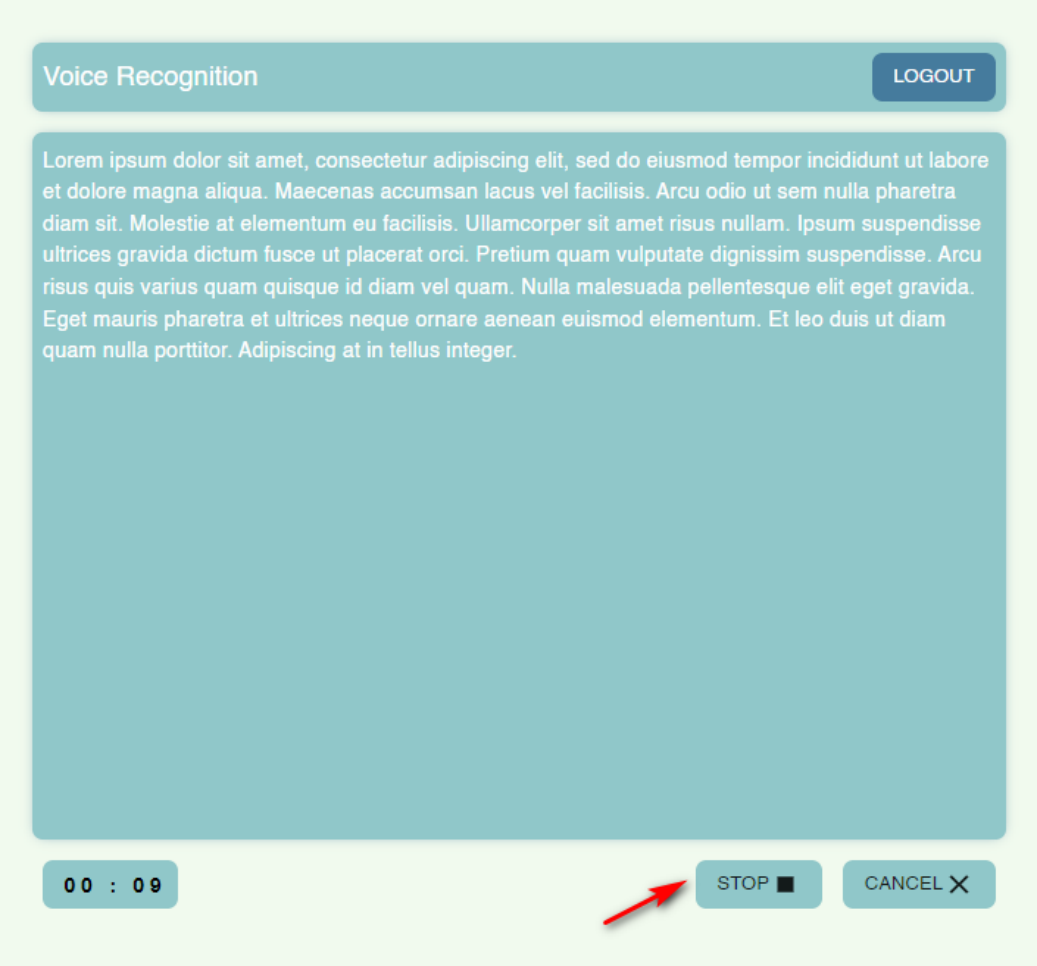
با انتخاب گزینه START در شکل ۳.۲ مرورگر از شما اجازه دسترسی به میکروفون دستگاه را می‌خواهد. پس از دادن دسترسی به مرورگر، گزینه STOP ظاهر می‌شود و ثانیه‌شمار شروع به شمارش می‌کند. پس از تکمیل تست شما باید گزینه STOP را کلیک کنید.

۵.۳.۲ ارسال صوت ضبط شده

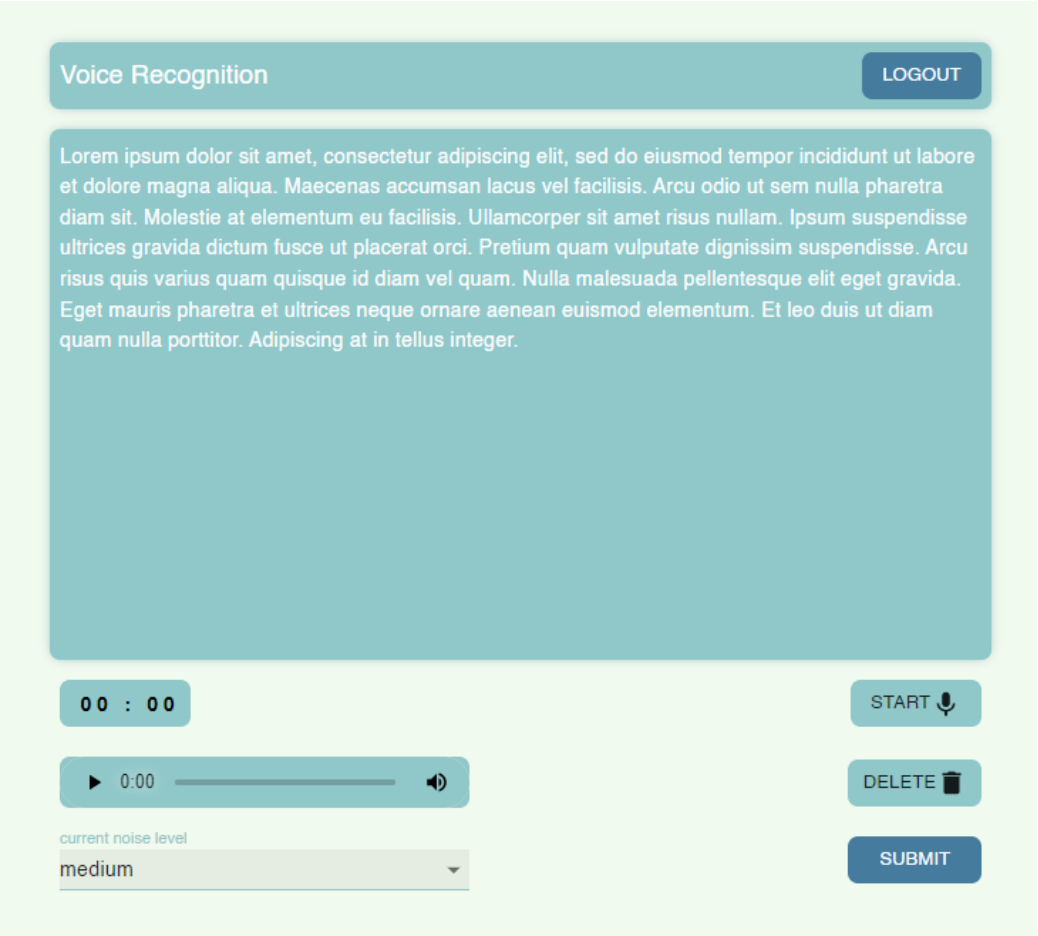
پس از انتخاب گزینه STOP در شکل ۴.۲ دکمه SUBMIT ظاهر می‌شود. اما قبل از آن شما باید میزان نویز فضای اطراف خود را بین ۳ گزینه نشان‌داده شده در منوی کشویی انتخاب کنید. با انتخاب گزینه SUBMIT شما تست را تکمیل کرده‌اید. همچنین می‌توانید با انتخاب گزینه DELETE صوت ضبط شده خود را پاک کنید و مجددا اقدام به ضبط صدای خود کنید.

۶.۳.۲ پنل مدیر سایت

این پنل به منظور مشاهده عملکرد کاربرها، مشاهده اطلاعات مربوط به آزمون‌ها و همچنین ایجاد آزمون‌های جدید برای کاربرها تعریف شده است. در تب ،USERS مدیر می‌تواند عملکرد هر کاربر و تعداد آزمون‌های انجام شده توسط هر کاربر را مشاهده کند. در تب ،TESTS مدیر می‌تواند مشاهده کند که هر آزمون توسط چه تعداد و چه کاربران بخصوصی انجام شده است. همچنین می‌تواند از طریق گزینه UPLOAD NEW TEST آزمون‌های جدیدی برای کاربران تعریف کند.

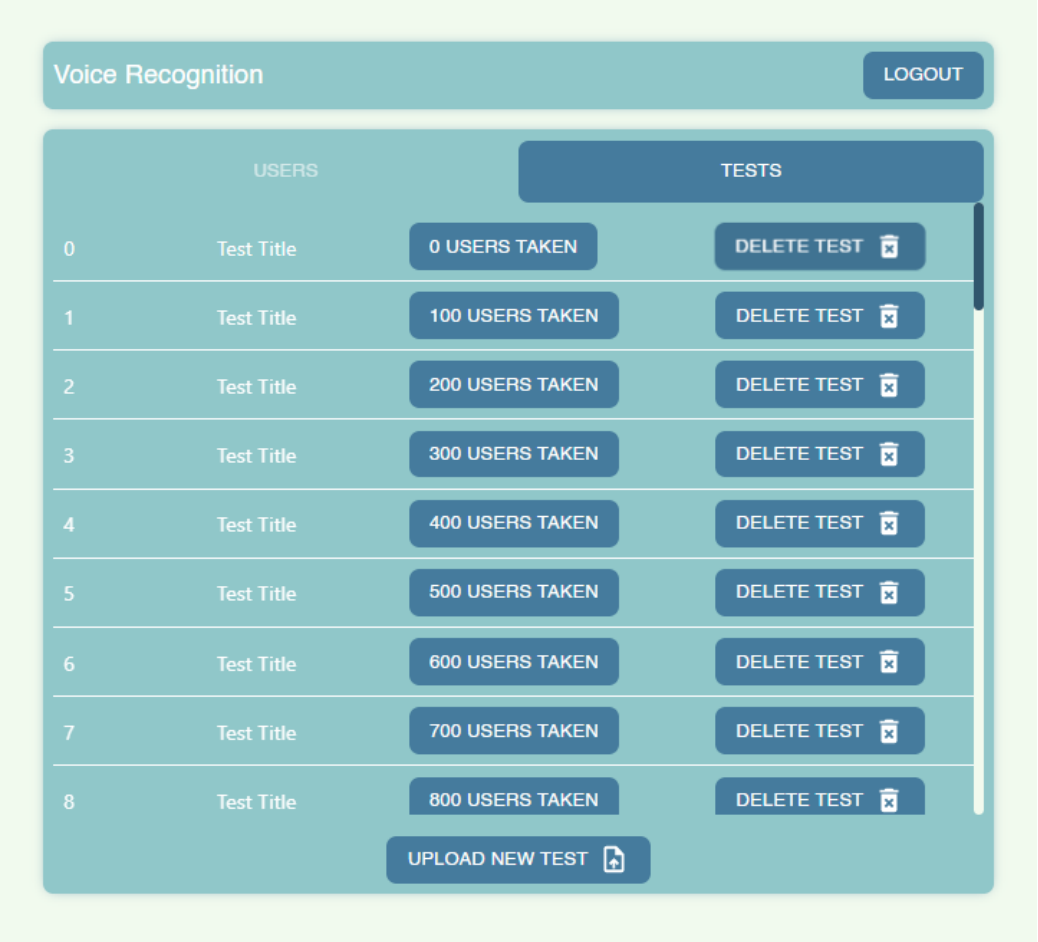


شکل ۴.۲: ضبط صدا



شکل ۵.۲: ارسال صوت ضبط شده

شکل ۶.۲: پنل ادمین - کاربران



شکل ۷.۲: پنل ادمین - آزمون‌ها

فصل ۳

چگونگی پیاده سازی پلتفرم

برای پیاده‌سازی این پلتفرم از ابزارهای گوناگونی استفاده شده است. از ابزارهای که مرتبط به ظاهر گرافیکی سایت هستند تا ابزارهایی که اتصال سیستم به سرور و پایگاه داده را ممکن می‌سازند.

۱.۳ ابزارهای Front-End

ابزارهای فرانت‌اند، منطق، ساختار، طراحی، رفتار و انیمیشن هر عنصری را که هنگام تعامل با وب‌سایت‌ها، برنامه‌های کاربردی وب و برنامه‌های تلفن همراه روی صفحه می‌بینید، تعیین می‌کنند. برای دستیابی به ظاهر مناسب و عملکرد مورد نظر ابزارهای متنوعی بکار گرفته شدند که این ابزارها عبارتند از:

- HTML, CSS, Java Script
- React.js
- Next.js

در ادامه به تفصیل در مورد هرکدام از این ابزارها و کاربردها آن‌ها در پیاده‌سازی وب اپلیکیشن می‌پردازیم.

۱.۱.۳ ابزارهای HTML، CSS و Java-Script

ما از این سه زبان برای طراحی ساختار، شکل دادن به ظاهر و افزودن عملکرد و کاربردهای مختلف به صفحات وب استفاده می‌کنیم. هنگامی که برخی از صفحات وب را با لینک‌ها، به همراه تمام دارایی‌های آن‌ها مانند تصاویر، ویدیوها و غیره که در رایانه سرور هستند به یکدیگر پیوند می‌دهید،

به یک وب‌سایت تبدیل می‌شود. این رندر معمولا در قسمت جلویی اتفاق می افتد، جایی که کاربران می‌توانند آنچه را که نمایش داده می‌شود ببینند و با آن تعامل داشته باشند.
از سوی دیگر، داده‌ها، به ویژه اطلاعات حساس مانند رمزهای عبور، از قسمت پشتی وب سایت ذخیره و عرضه می‌شوند. این بخشی از یک وب سایت است که فقط در رایانه سرور وجود دارد و در مرورگر جلویی نمایش داده نمی شود. در آنجا، کاربر نمی تواند آن اطلاعات را ببیند یا به راحتی به آن دسترسی پیدا کند. در بخش‌های بعدی در مورد آن صحبت خواهیم کرد.
به طور خلاصه، به عنوان یک توسعه دهنده وب، سه زبان اصلی که برای ساخت وب سایت استفاده می کنیم عبارتند از HTML[۱]، CSS[۲] و جاوا اسکریپت. جاوا اسکریپت زبان برنامه نویسی است که از برای تولید و ساخت ویژگی‌های عملکردی متفاوت در سایت استفاده می‌کنیم، از HTML برای ساختار سایت استفاده می‌کنیم و از CSS برای طراحی ظاهر و چیدمان صفحه وب استفاده می‌کنیم.
برای پیاده سازی این ۳ زبان باهم، ما از ابزار React.js استفاده کرده ایم. در بخش بعدی به توضیح این کتابخانه می‌پردازیم.

۲.۱.۳ ابزار React.JS

چارچوب React.js کتابخانه منبع باز جاوا اسکریپت است که توسط فیس بوک توسعه یافته است. این ابزار برای ساختن رابط‌های کاربری تعاملی و برنامه‌های کاربردی وب با سرعت بیشتر و کارآمدتر با میزان کد بسیار کمتری نسبت به جاوا اسکریپت استفاده می‌شود. در React، برنامه‌های خود را با ایجاد اجزای قابل استفاده مجدد توسعه می‌دهید که می‌توانید آنها را به عنوان بلوک‌های مستقل لگو در نظر بگیرید. این مؤلفه‌ها تکه‌های جداگانه یک رابط نهایی هستند که وقتی مونتاژ می‌شوند، کل رابط کاربری برنامه را تشکیل می‌دهند. نقش اصلی React در یک برنامه کاربردی این است که با ارائه بهترین و کارآمدترین اجرای رندر، لایه نمای آن برنامه را درست مانند V در الگوی مدل_view_کنترلر (MVC) مدیریت کند. React.js به‌جای پرداختن به کل رابط کاربری به‌عنوان یک واحد مستقل، توسعه‌دهندگان را تشویق می‌کند تا این رابط‌های کاربری پیچیده را به اجزای منفرد قابل استفاده مجدد که بلوک‌های ساختمان کل رابط کاربری را تشکیل می‌دهند، جدا کنند. در انجام این کار، چارچوب ReactJS سرعت و کارایی جاوا اسکریپت را با یک روش کارآمدتر برای دستکاری DOM[۳] ترکیب می‌کند تا صفحات وب را سریع‌تر ارائه کند و برنامه‌های وب بسیار پویا و پاسخگو ایجاد کند.
پایه و اساس کدهای داخل پروژه برای ساخت این پلتفرم react.js است. استفاده از این کتابخانه به ما انعطاف بسیار زیادی برای پیاده سازی ویژگی‌های مورد نیازمان می‌دهد. این کتابخانه شامل فریم ورک‌هایی مانند view.js و next.js برای استفاده است. ما از فریم ورک next.js در پروژه بهره بردیم که در بخش بعدی راجع به این ابزار قدرتمند صحبت می‌کنیم.

---
[۱] Hyper Text Markup Language
[۲] Cascading Style Sheets
[۳] Document Object Model

چارچوب MUI

MUI یک کتابخانه عظیم از اجزای UI است که طراحان و توسعه دهندگان می‌توانند از آن برای ساخت برنامه‌های React استفاده کنند. این پروژه متن باز از دستورالعمل‌های Google برای ایجاد مؤلفه‌ها پیروی می‌کند و یک کتابخانه قابل تنظیم از عناصر اساسی و پیشرفته UI در اختیار شما قرار می‌دهد. MUI همچنین مجموعه‌ای از قالب‌ها و ابزارهای React را به فروش می‌رساند و رابط‌های کاربری آماده‌ای را در اختیار شما قرار می‌دهد تا پروژه خود را تغییر دهید. طراحان اغلب از کیت‌های UI برای ساخت محصولات جدید یا ویژگی‌های افزودنی برای پروژه‌های موجود استفاده می‌کنند. این کتابخانه ها به طراحان اجازه می‌دهد تا اجزای مورد نیاز خود را برای طراحی سریع رابط ها بکشند و رها کنند.

ما در هنگام پیاده سازی این پروژه، برای طراحی ظاهر و سفارشی سازی رابط کاربری صفحات پلتفرم از این چارچوب بهره بردیم تا آزادی عمل بیشتر و سرعت بیشتری در مسیر پیاده‌سازی پلتفرم داشته باشیم.

State Management در React

فرایند React State management برای مدیریت داده‌هایی است که مؤلفه‌های React برای رندر کردن خود به آن‌ها نیاز دارند. این داده‌ها معمولاً در object state مؤلفه ذخیره می‌شوند. هنگامی که state یک object تغییر می‌کند، کامپوننت خود را دوباره رندر می‌کند. اساساً نیمی از یک برنامه React فرایند React State management است و شامل تمام داده‌ها می‌شود. نیمه دیگر ارائه شامل HTML، CSS و قالب‌بندی است. ابزارهای متنوعی در React برای این فرایند وجوددارد. یکی از این ابزارها Redux است. از آنجایی که مدیریت حالت‌ها در این پروژه پیچیده نیست، ما از استفاده Redux در پروژه خود صرف نظر کردیم چرا که علی‌رغم کارایی بالا، استفاده از این ابزار موجب سنگین شدن بیش از حد کدهای پروژه می‌شود. انتخاب ما استفاده از State management خود React و بدون بکارگیری از کتابخانه یا فریم ورک اضافی به همراه استفاده از local storage برای نگهداری داده‌ها است. local storage مربوط به cache مرورگر می‌شود. به طور مثال، هنگام ورود یک کاربر به حساب کاربری، اطلاعات آن کاربر در این فضا ذخیره شده تا از صدا زدن چندباره هاAPI جلوگیری شود.

۳.۱.۳ ابزار Next.JS

Next.js یک چارچوب توسعه وب متن باز است که توسط Vercel ایجاد شده است و برنامه‌های وب مبتنی بر React را با رندر سمت سرور و تولید وب‌سایت‌های ثابت امکان‌پذیر می‌کند. مستندات React از Next.js در میان «زابزارهای توصیه‌شده» نام می‌برد که به توسعه‌دهندگان به عنوان راه‌حلی در هنگام «ساخت یک وب‌سایت ارائه‌شده توسط سرور با «Node.js توصیه می‌کند.[۱۹] در جایی که برنامه‌های سنتی React فقط می‌توانند محتوای خود را در مرورگر سمت کلاینت ارائه

دهند، Next.js این قابلیت را گسترش می‌دهد تا برنامه‌های ارائه‌شده در سمت سرور را نیز در بر بگیرد.
حق چاپ و علائم تجاری Next.js متعلق به Vercel است، [۲۰] که همچنین توسعه منبع باز آن را حفظ و رهبری می کند.[۲۱]

۲.۳ ابزارهای Back-End

ابزارهای ،Backend کتابخانه‌هایی از زبان‌های برنامه‌نویسی سمت سرور هستند که به ساختن ساختار باطن یک وب‌سایت کمک می‌کنند. ابزارهای Backend اجزای آماده‌ای را برای توسعه یک وب اپلیکیشن پویا ارائه می‌دهند. استفاده از این فریم‌ورک‌ها تیاز به ساخت و پیکربندی همه چیز از ابتدا را حذف می‌کنند و به توسعه‌دهندگان یک شروع می‌دهند. ابزارهایی که در این پروژه استفاده شدند عبارتند از:

- MySQL
- REST API

۱.۲.۳ ابزار MySQL

MySQL یک سیستم مدیریت پایگاه داده رابطه‌ای متن باز (RDBMS) با مدل کلاینت_سرور است. RDBMS نرم افزار یا سرویسی است که برای ایجاد و مدیریت پایگاه‌های داده بر اساس مدل رابطه ای استفاده می شود. برای ساخت جداول دیتابیس و قراردادن اطلاعات یوزرها و تسک‌ها و مدیریت آن‌ها از این ابزار در پروژه استفاده شده است.

۲.۲.۳ ابزار REST API

یک API[۴]، مجموعه‌ای از قوانین است که نحوه اتصال و ارتباط برنامه‌ها یا دستگاه‌ها با یکدیگر را مشخص می کند. REST API یک API است که با اصول طراحی REST یا representational state transfer architectural style مطابقت دارد. به همین دلیل گاهی اوقات این ابزار RESTful API نیز گفته می‌شود.
یک API از طریق درخواست‌های HTTP برای انجام عملکردهای استاندارد پایگاه داده مانند ایجاد، خواندن، به روز رسانی و حذف رکوردها (همچنین به عنوان CRUD شناخته می‌شود) در یک منبع ارتباط برقرار می‌کنند. به عنوان مثال، یک REST API از یک درخواست GET برای بازیابی یک رکورد، یک درخواست POST برای ایجاد یک رکورد، یک درخواست PUT برای به

[۴] Application Programming Interface

روزرسانی یک رکورد و یک درخواست DELETE برای حذف یک رکورد استفاده می‌کند. همه روش‌های HTTP را می‌توان در تماس‌های API استفاده کرد. یک REST API که به خوبی طراحی شده است شبیه به وب سایتی است که در یک مرورگر وب با قابلیت HTTP داخلی اجرا می‌شود.

ابزار Axios

ابزار Axios یک سرویس گیرنده HTTP برای جاوا اسکریپت است. این ابزار توانایی ایجاد درخواست‌های HTTP از مرورگر و مدیریت تبدیل داده‌های درخواست و پاسخ به آن‌ها را دارد. در این پروژه برای وصل شدن به API ها از این ابزار استفاده شده است.

کتاب‌نامه

[۱] Calado, A., Freitas, J., Braga, D., Dias, M., “Multi-Language Telephony Speech Data Collection and Annotation”. in: Braga et al. (eds.) Propor 2008 Special Session: Applications of Portuguese Speech and Language Technologies, September 10, 2008, Curia, Portugal, 2008.

[۲] Neto, N., Patrick, C., Adami, A.G., Klautau, A. G., “Spoltech and ogi-22 baseline systems for speech recognition in Brazilian portuguese”, in Teixeira, A., Lima, V., Oliveira, L., Quaresma, P. (eds) Propor 2008, LNCS (LNAI), vol. 5190, pp. 256-259, Springer, Heidelberg, 2008.

[۳] Freitas, J., Calado, A., Braga, D., Silva, P., & Dias, M. (2010). Crowdsourcing platform for large-scale speech data collection. Proc. Fala.

[۴] D. A. Reynolds, ”An overview of automatic speaker recognition technology”, Proc. IEEE Int. Conf. Acoust. Speech Signal Process., vol. 4, pp. 4072-4075, May 2002.

[۵] S. Furui, ”Speaker-dependent-feature extraction recognition and processing techniques”, Speech Commun., vol. 10, no. 5, pp. 505-520, Dec. 1991.

[۶] G. Zhou, J. H. L. Hansen and J. F. Kaiser, ”Nonlinear feature based classification of speech under stress”, IEEE Trans. Speech Audio Process., vol. 9, no. 3, pp. 201-216, Mar. 2001.

[۷] C. Fredouille, G. Pouchoulin, J.-F. Bonastre, M. Azzarello, A. Giovanni and A. Ghio, ”Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia)”, Proc. Eur. Conf. Speech Commun. Technol. (Eurospeech), pp. 149-152, 2005.

[۸] V. A. Petrushin, "Emotion recognition in speech signal: Experimental study development and application", Proc. 6th Int. Conf. Spoken Lang. Process. (ICSLP), pp. 5, 2000.

[۹] V. A. Petrushin, "Emotion recognition in speech signal: Experimental study development and application", Proc. 6th Int. Conf. Spoken Lang. Process. (ICSLP), pp. 5, 2000.

[۱۰] E. Nöth, S. Harbeck and H. Niemann, "Multilingual speech recognition" in Computational Models of Speech Pattern Processing, Springer, vol. 169, pp. 362-374, 1999.

[۱۱] T. Schultz and A. Waibel, "Language-independent and language-adaptive acoustic modeling for speech recognition", Speech Commun., vol. 35, no. 1, pp. 31-51, 2001.

[۱۲] C. Chelba, T. J. Hazen and M. Saraclar, "Retrieval and browsing of spoken content", IEEE Signal Process. Mag., vol. 25, no. 3, pp. 39-49, May 2008.

[۱۳] F. Biadsy, "Automatic dialect and accent recognition and its application to speech recognition", pp. 1-171, 2011.

[۱۴] T. Bocklet, A. Maier, J. G. Bauer, F. Burkhardt and E. Nöth, "Age and gender recognition for telephone applications based on GMM supervectors and support vector machines", Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 1605-1608, Mar./Apr. 2008.

[۱۵] T. Vogt and E. André, "Improving automatic emotion recognition from speech via gender differentiation", Proc. Lang. Resour. Eval. Conf., pp. 1123-1126, Jan. 2006.

[۱۶] I. M. A. Shahin, "Gender-dependent emotion recognition based on HMMs and SPHMMs", Int. J. Speech Technol., vol. 16, no. 2, pp. 133-141, 2013.

[۱۷] Song, Z. English speech recognition based on deep learning with multiple features. Computing, no. 102(3), pp. 663-682, 2020.

[۱۸] Reed. L, The Requirements and Applications of Speech Recognition Technology for Voice Activated Command and Control in the Tactical Military Environment.Army Communications-Electronics Command Fort Monmouth NJ. 2004

[۱۹] Vryzas N., Kotsakis R., Liatsou A., Dimoulas C., Kalliris G. Speech emotion recognition for performance interaction Journal of the Audio Engineering Society, no. 66 (6) (2018), pp. 457-467

[۲۰] Recommended Toolchains, (HTML). React documentation. 2023

[۲۱] Next.js Brand Guidelines, 26 August 2022

[۲۲] Develop. Preview. Ship. For the best frontend teams – Vercel, (HTML). vercel.com. Archived from the original on 10 July 2022

Abstract

One of the basic requirements for the development of a speech recognition and processing system is to have speech data with appropriate volume and quality. Speech datasets should have sufficient diversity in terms of age, gender, level of education and accent of the speakers. In this project, the goal is to design and implement a web-based platform to collect audio data in bulk from different speakers. This platform will have the ability to be configured for different speech collection scenarios with different applications and will have a user-friendly graphical interface.

College of Science
School of Mathematics, Statistics, and Computer Science

Design and implementation of a web-based platform for audio data collection

Amirmasoud Aghaei

Supervisor: Dr. Bagher Babaali

A thesis submitted in partial fulfillment of the requirements for
the degree of B.Sc. in Computer Science

Februray 2023