بهبود احراز هویت مبتنی بر ریسک با ترکیب ویژگی‌های جعل خوبی

الهام یزدانی شهربابک[۱]، مهدیه قزوینی[۲]، زهرا کریمی دهکردی[۳]

۱-دانشجوی کارشناسی ارشد مهندسی کامپیوتر، بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان، ایران، ایران

۲-عضو هیأت علمی بخش مهندسی کامپیوتر، دانشگاه شهید باهنر کرمان، ایران، ایران.mghazvini@uk.ac.ir

۳-عضو هیات علمی گروه مهندسی کامپیوتر، دانشگاه شهرکرد، شهرکرد، ایران. Zahra.karimi@sku.ac.ir

چکیده

احراز هویت به فرایندی گفته می‌شود که طی آن درستی هویت یک فرد (حتی یک شیء) تأیید می‌گردد. احراز هویت مبتنی بر ریسک یکی از انواع احراز هویت است که بر مبنای میزان ریسکی که از رفتارهای غیرعادی و داده‌های غیرعادی کاربر به دست می‌آورد سعی در احراز هویت کاربر می‌کند. در این پژوهش هدف بررسی احراز هویت مبتنی بر ریسک با لحاظ‌کردن رفتارهای جعلی است. افراد جاعل ممکن است به زمان بیشتری برای پاسخ به سؤالات شریک گفتگو نیاز داشته باشند، یا ممکن است از اصلاحاتی برای تغییر متن خود استفاده کنند تا از تناقض در پاسخ‌هایشان جلوگیری شود. این مسائل باعث ایجاد تفاوت‌هایی در رفتار تایپ می‌شود که می‌توان آن را در ضرب آهنگ تایپ اندازه‌گیری کرد. نتایج نشان داد که ترکیب جعل خوبی با ویژگی «سرعت فشردن کلید»، نسبت به سایر ویژگی‌ها، بهبود بیشتری در احراز هویت را نتیجه می‌دهد.

کلمات کلیدی: احراز هویت، اصالت، ریسک، مالکیت، دینامیک موشواره، دست خط صفحه‌کلید.

۱- مقدمه

شناسایی مشتریان یا احراز هویت (Authentication) فرایندی است که بر اساس آن بررسی می‌شود که آیا طرف مقابل ارتباط، همانی است که باید باشد یا یک نفوذگر است که خود را به‌جای طرف واقعی جا زده است. احراز هویت مبتنی بر ریسک (RBA) یک سیستم احراز هویت پویا بدون اجبار کاربران به استفاده از احراز هویت چند فاکتوری است. در این روش، مشخصات زمینه‌ای اتصال کاربر باید شناسایی و برای محاسبه شاخص ریسک (index risk) استفاده شود. در راه‌اندازی احراز هویت مبتنی بر ریسک، اغلب از احراز هویت ماشین استفاده می‌شود. احراز هویت ماشین در پس‌زمینه اجرا می‌شود و فقط درصورتی‌که رایانه شناسایی نشود از مشتری درخواست می‌شود اطلاعات اضافی برای تأیید اعتبار وارد کند .در این روش بدون ایجاد مزاحمت برای کاربر دقت اعتبارسنجی بهبود می‌یابد. در حال حاضر بیشتر سیستم‌های RBA بر اطلاعات اولیه ارتباطات وب مانند آدرس IP منبع یا سرعت تراکنش‌های انجام‌شده توسط یک حساب خاص یا منشأ آن از یک آدرس IP خاص تکیه می‌کنند. چنین اطلاعاتی را می‌توان به‌راحتی جعل کرد و به‌این‌ترتیب، استحکام و قابلیت اطمینان سیستم‌های پیشنهادی کاهش می‌یابد.

ازاین‌رو در سیستم‌های آنلاین RBA پیشنهاد شده است که با ترکیب دینامیک ماوس و صفحه‌کلید (دستخط صفحه‌کلید) اطلاعات هویتی قوی‌تری از کاربر مدنظر قرار گیرد.

دینامیک ماوس، ویژگی‌های حرکت ماوس در هنگام کار با واسط کاربر گرافیکی و دینامیک ضربه کلید نیز شامل اطلاعاتی مانند زمان ضربه یک کلید و فاصله زمانی بین دو ضربه است که یک بیومتریک رفتاری مؤثر برای احراز هویت کاربر در ترمینال کامپیوتر محسوب می‌شوند. اخیراً احراز هویت مداوم یا فعال با استفاده از دینامیک ضربه‌زدن به کلید علاقه زیادی را در بین محققان ایجاد کرده است. رفتار تایپ برای اطلاعاتی که معمولاً در فرم‌های ثبت‌نام آنلاین موردنیاز است (به‌عنوان‌مثال، نام، نام خانوادگی و ایمیل)، به طور قابل‌توجهی بین پاسخ‌های صادقانه یا غیرواقعی متفاوت است.

تجزیه‌وتحلیل ضربه‌زدن به کلید می‌تواند پتانسیل زیادی در تشخیص صحت اطلاعات اعلام شده داشته باشد و می‌تواند برای تعداد زیادی از سناریوهای عملی که کاربران را ملزم به واردکردن داده‌های شخصی از راه دور از طریق صفحه‌کلید می‌کنند، اعمال شود. فریبکاری به جعل هویت خلاصه نمی‌شود. افراد ممکن است در جایگاه و هویت خودشان آگاهانه و یا ناآگاهانه با ارائه اطلاعات نادرست و غیرواقعی، روایت‌های فریبنده تولید کنند. تولید فریب یک فرایند روانی پیچیده است و اطلاعات کمی در مورد تولید فریب، و اینکه چگونه مکانیسم‌های آن ممکن است به تمایز روایت‌های واقعی از فریبکارانه کمک کند، در دست است .فریبنده بودن از نظر شناختی پیچیده‌تر از گفتن حقیقت است و این پیچیدگی بالاتر در تغییرات رفتار کاربر در طول یک کار منعکس می‌شود به‌عبارت‌دیگر، ثبت تغییرات رفتاری برای پی‌بردن به اینکه آیا کاربری که وظیفه‌ای را انجام می‌دهد (مثال به سؤالات پاسخ می‌دهد) دروغ می‌گوید یا خیر، امکان‌پذیر است

در بخش دوم مختصری ازکارهای مزتبط امده است. بخش سوم به ارائه روش پیشنهادی می پردازد. نتایج و ارزیابی در بخش چهارم آمده و درن هایت در بخش پنجم مقاله با جمع بندی و نتیجه گیری به پیایان می‌رسد.

۲- کارهای مرتبط

در زمینه احراز هویت کاربران مبتنی بر ریسک روش‌های مختلفی ارائه شده است. در [۱] ضربات صفحه‌کلید را به‌عنوان ابزاری برای دسترسی به فرایند نوشتن بلادرنگ نویسندگان آنلاین در زمینه تشخیص فریب معرفی نموده است. نشان‌داده‌شده که تفاوت در الگوهای ضربه‌زدن به کلید مانند مانورهای ویرایش و مدت مکث می‌تواند به تمایز بین نوشته‌های واقعی و فریبنده کمک کند. نتایج تجربی نشان می‌دهد که ترکیب ویژگی‌های مبتنی بر ضربه کلید منجر به بهبود عملکرد در تشخیص فریب می‌شود. تلاش‌های عمدی برای نشان‌دادن خود به شیوه‌های غیرواقعی معمولاً در مدیریت شخصیت رخ می‌دهد. پرسش‌نامه‌ها هدف اصلی مطالعه [۲] با بررسی این موضوع بود که آیا شاخص‌های زمانی ردیابی ماوس و مدل‌های یادگیری ماشینی می‌توانند تشخیص سوژه‌هایی را که سبک پاسخ جعلی-خوب را هنگام پاسخ‌دادن به فهرست‌های شخصیتی با چهار گزینه، با و بدون فشار زمان، اجرا می‌کنند، بهبود بخشند. تکنیک‌های جدید اساساً مبتنی بر استراتژی‌های شناختی هستند که می‌توانند بارهای شناختی دروغگوها را افزایش دهند و با اندازه‌گیری‌های RT و پویایی ضربه‌زدن به ماوس یا کلید ترکیب می‌شوند.

در الگوریتم پیشنهادی [۳] بر روی دو ویژگی قابل‌اندازه‌گیری تمرکز شده است، دینامیک اثر انگشت و ضربه‌زدن به کلید. در طول آزمایش‌ها، از رویکردهای متنوعی مانند-k نزدیک‌ترین همسایه، means-k، طبقه‌بندی Bayes Naive و درخت‌های تصمیم استفاده کرده‌اند. تجزیه‌وتحلیل انجام شده نشان داده است که به‌وضوح می‌توان هویت انسان را بر اساس پویایی ضربه‌زدن به کلید با سطح دقت رضایت‌بخشی تشخیص داد. علاوه بر این، مشخص شده که انتخاب الگوریتم یادگیری ماشین تأثیر زیادی بر کیفیت طبقه‌بندی و تشخیص دارد [۳].

ما قصد داریم احراز هویت مبتنی بر ریسک را مورد بررسی جامع قرار بدهیم. احراز هویت مبتنی بر ریسک یکی از انواع احراز هویت است که بر مبنای میزان ریسکی که از رفتارهای غیرعادی و داده‌های غیرعادی کاربر به دست می‌آورد سعی در

احراز هویت کاربر می‌کند، این اطلاعات معمولاً شامل شناسه کاربر، گذرواژه، مکان، لاگین زمینه‌های استفاده، شیوه تایپ‌کردن کاربر و... هر چه میزان این ریسک بالاتر باشد سیستم از چالش‌های قوی‌تری برای احراز هویت کاربر استفاده می‌کند.

۳- روش پیشنهادی

۳-۱- معرفی دیتاست های مورد استفاده

۳-۱-۱- معرفی دیتاست احراز هویت [۴] استفاده‌شده در پژوهش جاری

این دیتاست، مجموعه‌ای از داده‌های بیومتریک رفتاری را ارائه می‌کند که معمولاً به‌عنوان دینامیک صفحه‌کلید و ماوس شناخته می‌شود. مجموعه‌داده شامل ۱۷۶۰ نمونه از داده‌های بیومتریک است که در ۸۸ جلسه و هر جلسه با یک کاربر متمایز با برنامه گرافیکی جمع‌آوری شده است. در هر جلسه، کاربر ۲۰ بار داده وارد کرده است. به کاربر اطلاعات یک کارت ساختگی (انتخاب‌شده به‌صورت تصادفی از یک مجموعه کارت) اختصاص‌داده‌شده تا ۱۰ بار وارد شود و متعاقباً ۱۰ کارت واقعی که هر یک باید یک‌بار وارد می‌شده است؛ بنابراین، در مجموع ۲۰ نمونه داده (یعنی ۱۰ نمونه قانونی و ۱۰ غیرقانونی) در طول هر جلسه ورود کاربر در برنامه جمع‌آوری‌شده است. داده‌ها به‌صورت ۸۸ فایل اکسل ارائه شدند که در هر فایل ۲۰ داده با ویژگی‌های زیر وجود دارند:

۱- متوسط زمان dwelling: زمان dwelling زمان بین فشردن یک کلید تا رهاکردن آن است
۲- متوسط زمان flight: زمان flight زمان بین رهاکردن یک کلید تا فشردن کلید بعدی است
۳- متوسط مسیر: مسیری که کاربر با انگشت یا ماوس طی کرده است
٤- برچسب: دارای دو مقدار ۱ (قانونی) و ۰ (غیرقانونی) است.

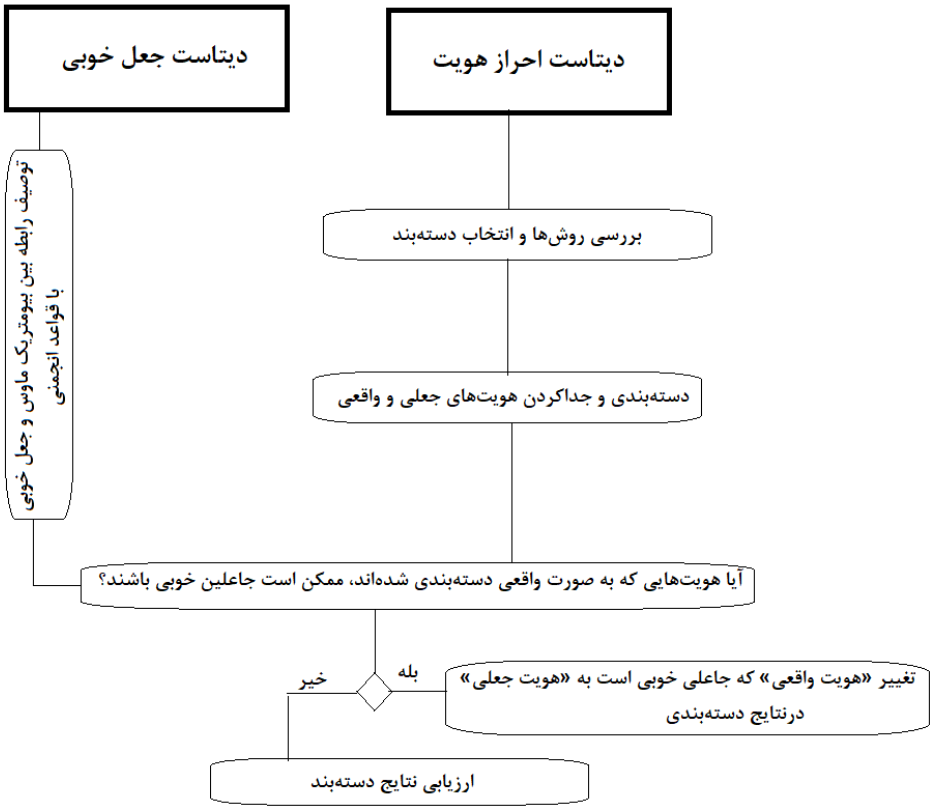
۳-۱-۲-معرفی دیتاست جعل خوبی[۵] استفاده‌شده در پژوهش

در پژوهش[5]، تعداد ۱۲۰ داوطلب ۹۶ مورد از تست‌های شخصیت L، K و S MMPI-2؛ و VR و مقیاس PPI-R را کامل کردند. یک گروه ابتدا با دستورالعمل پاسخ صادقانه (الف) و سپس با دستورالعمل پاسخ جعلی (ب) تست‌ها را جواب دادند و یک گروه دیگر بر عکس ابتدا با دستورالعمل پاسخ جعلی (الف) و سپس با دستورالعمل پاسخ صادقانه (ب) تست‌ها را جواب دادند. از اسکرین شرکت‌کنندگان در هنگام پاسخ‌گویی فیلم‌برداری شد و ویژگی‌های زیر جمع‌آوری شد

- ماکزیمم انحراف (maximum deviation): بیشترین فاصله از مسیر ایدئال
- تفاضل ناحیه زیر نمودار در مسیر طی شده نسبت به مسیر ایده‌آل (AUC)
- زمان پاسخ: زمان بین ظاهرشدن سؤال روی صفحه‌نمایش و کلیک‌کردن روی دکمه شروع
- زمان رسیدن به نقطه ماکزیمم (MAD Time) زمانی که طول می‌کشد تا کاربر به نقطه ماکزیمم انحراف که در شکل ۴-۲ نیز مشخص شده است، برسد.
- سرعت ماوس نسبت به محور افقی (vel_x)
- سرعت ماوس نسبت به محور عمودی (vel_y)

۲-۳-روش پیشنهادی احراز هویت مبتنی بر ریسک با ترکیب ویژگی‌های جعل خوبی

به‌منظور ترکیب انجام پژوهش با کمک‌گرفتن از دیتاست جعل خوبی یک صفت «احتمال جعل خوبی» به دیتاست احراز هویت اضافه شد و میزان بهترشدن کارایی دسته‌بندی تحلیل شد. مراحل روش در شکل ۱ نشان داده شده‌اند.



شکل ۱ فلوچارت روش پیشنهادی

۱-۲-۳-حذف داده‌های پرت

برای یافتن داده‌های پرت تک‌ستونی از دامنه میان چارکی (IQR) استفاده شد.

در این جا، مقادیری که فاصله آنها از چارک سوم، بیشتر از سه برابر دامنه میان چارکی باشد، نقاط پرت محسوب شدند و همچنین مقادیری که از چارک اول، بیشتر از سه برابر دامنه میان چارکی کوچک‌تر بودند نیز نقاط پرت محسوب شدند سطرهایی که یکی از ویژگی‌های بیومتریک رفتاری آنها، پرت بود، از دیتاست حذف شدند. به این منظور از فیلتر InterQuartileRange استفاده شد. برای یافتن داده‌های پرت چند ستونی از فاصله ماهالونوبیس استفاده شد. به‌این‌ترتیب احتیاج به اندازه‌گیری فاصله یک نقطه از یک توزیع داریم. فاصله اقلیدسی به این منظور مناسب نیست. زیرا اولاً چنانچه ویژگی‌ها با هم همبستگی داشته باشند، فاصله اقلیدسی بزرگ‌تر از حالتی است که مستقل باشند. به‌این‌ترتیب به نظر می‌رسد که فاصله اقلیدسی وزن بیشتری به مشاهدات مرتبط با هم می‌دهد. ثانیاً اگر واقعاً شامل نقاط پرت باشد، میانگین و انحراف استاندارد را تحت‌تأثیر خود قرار داده و محاسبات را دچار مشکل می‌کنند. برای خروج از این بحران از فاصله ماهالانوبیس استفاده می‌کنیم. در نهایت دیتاست جعل خوبی، دو داده پرت حذف شدند و تعداد ۲۳۹ سطر باقی ماندند. در دیتاست احراز هویت، ۵۸ سطر حذف شدند و تعداد ۱۷۰۲ سطر باقی ماندند.

۲-۲-۳- بررسی روش‌ها و انتخاب دسته‌بند

جدول ۱ نتایج دسته‌بندی روش‌های مختلف در وکا

روش	درصد مثبت واقعی	درصد منفی واقعی
نایبوبیز	۰/۵۷۵	۰/۴۳۹
نایبوبیز مالتی نومیال	۰/۷۲۵	۰/۲۶۴
لاجستیک	۰/۴۷۳	۰/۵۷۹
مالتی لیر پرسپترون	۰/۵۰۶	۰/۵۱۹
SGD	۰/۳۸۱	۰/۶۴۱
Svm	۰/۳۹۳	۰/۶۱۲
پرسپترون رای‌گیرنده	۰/۸۰۱	۰/۲۰۱
IBK	۰/۵۵۳	۰/۵۴۰
LWL	۰/۶۱۹	۰/۴۴۹
CalassificationviaRegression	۰/۶۱۵	۰/۴۰۴
Iterative classifier optimizer	۰/۲۳۹	۰/۷۴۱
J48	۰/۹۳۷	۰/۰۴۳
جنگل تصادفی	۰/۵۳۳	۰/۴۴۷

به‌منظور انجام بررسی، روش‌های مختلف در وکا دسته‌بندی شدند. در این نتایج، درصد مثبت واقعی (هویت‌های جعلی که به‌صورت جعلی دسته‌بندی شدند) که همان یادآوری[1]، است گزارش شده است. حداکثر مقدار این معیار یک و یا ۱۰۰ درصد و حداقل مقدار آن صفر است و هرچه مواردی که ما انتظار داشتیم پیش‌بینی شوند؛ ولی برنامه پیش‌بینی نکرده است نسبت به‌پیش بینی‌های درست بیشتر باشد مقدار معیار یادآوری کمتر خواهد شد. طبیعی است که با افزایش درصد مثبت واقعی، درصد منفی واقعی (هویت‌های واقعی که به‌صورت واقعی دسته‌بندی شدند) کاهش می‌یابد؛ بنابراین درصد منفی واقعی نیز گزارش شده است. دسته‌بندی به‌صورت ۱۰ لایه‌ای و تقاطعی انجام شده است.

جدول ۲ نتایج دسته‌بندی در ۵۰ بار آزمایش ۱۰ لایه‌ای

	متوسط معیار F1	انحراف معیار F1
نایبوبیز مالتی نومیال	۰/۵۸۵۲	۰/۰۴۵۶
پرسپترون رای‌گیرنده	۰/۶۶۴۹	۰/۰۴۲۶
J48	۰/۶۴۹۰	۰/۰۵۴۵

چنانچه مشخص است دسته‌بندهای نایبوبیز مالتی نومیال ، پرسپترون رای‌گیرنده و j48 نسبت به سایر دسته‌بندها عملکرد بهتری دارد. جهت به‌دست‌آوردن نتایج دقیق‌تر در یک آزمایش با ۵۰ با تکرار این سه دسته‌بند بررسی شدند. نتایج این بررسی در جدول ۲ گزارش شده است. نتایج حاکی از این است که پروسپترون رای‌گیرنده بهترین نتایج را دارد.

۳-۲-۳- افزودن ویژگی‌های جعل خوبی به دیتاست احراز هویت

به‌ازای داده‌های بیومتریک رفتاری در هر سطر از دیتاست احرار هویت، میزان جعل خوبی با استفاده از دیتاست جعل خوبی تخمین زده می‌شود. به این شکل که دیتاست‌ها به بهترین شکل ممکن خوشه‌بندی می‌شوند. متوسط جعل خوبی در آن خوشه از دیتاست جعل خوبی به‌عنوان میزان جعل خوبی در خوشه متناظرش در دیتاست احراز هویت قرار می‌گیرد برای

[1] recall

خوشه‌بندی از روش k-means استفاده کردیم. قبل از اجرای این روش باید بهترین مقدار k را پیدا کنیم. بهترین خوشه‌بندی، خوشه‌بندی است که در آن فاصله نقاط یک خوشه به همدیگر کمترین مقدار ممکن و درعین‌حال فاصله نقاط خوشه‌های مختلف از همدیگر بیشترین باشد. به این منظور از شاخص دیویس بولدین استفاده شد. شاخص دیویس بولدین برای تمام مقادیر k بین ۲ تا ۱۱ به‌دست‌آمده است. مشاهده شد شاخص دیویس بولدین برای k=3 از سایرین کمتر است؛ بنابراین بهترین خوشه‌بندی همان خوشه‌بندی سه‌گانه است. آمار مربوط به خوشه‌بندی سه‌گانه دیتاست جعل خوبی در ادامه گزارش شده است.

جدول ۳ تعداد نمونه‌ها و متوسط ویژگی‌ها در هر خوشه در دیتاست جعل خوبی

	تعداد	درصد	mad	auc	rt	madt	velx	vely	جعل خوبی
خوشه یک	۱۲۰	۵۱	۰/۵۱۳	۱/۱۱۲	۳۱۶۶	۱۸۲۱	-۰/۰۰۱۱	۰/۰۱۳۰	۰/۴۰۸۳
خوشه دو	۸۲	۳۵	۰/۵۳۰	۱/۱۷۲	۴۸۱۲	۲۷۶۹	۰/۰۰۰۱	۰/۰۱۳۰	۰/۵۴۸۸
خوشه سه	۳۴	۱۴	۰/۴۸۸	۱/۱۳۶	۷۱۴۴	۴۴۱۵	۰/۰۰۱۰	۰/۰۱۳۰	۰/۷۰۵۶

چنانچه در جدول ۳ مشخص است، احتمال جعل خوبی در خوشه سه از سایر خوشه‌ها بیشتر است. در این خوشه زمان واکنش (rt) و زمان رسیدن به ماکزیمم انحراف (madt) نیز نسبت به دو خوشه دیگر بیشتر است. یعنی پاسخ‌دهندگان این خوشه در هنگام شروع یک کار، کندتر واکنش نشان دادند. در خوشه یک، احتمال جعل خوبی نسبت به سایرین کمتر است و تفاوت ناحیه زیر نمودار (auc) و فاصله تا ماکزیمم انحراف (mad) نیز از سایر پاسخ‌دهندگان کمتر است. به این معنی که در این خوشه، پاسخ‌دهندگان مسیر کوتاه‌تری طی کردند. در خوشه دو احتمال جعل خوبی حدود متوسط است و تفاوت ناحیه زیر نمودار (auc) و فاصله تا ماکزیمم انحراف (mad) نیز از سایر پاسخ‌دهندگان بیشتر است؛ ولی سرعت ماوس velx در این خوشه نسبت به این مقدار در خوشه‌های دیگر کمتر است. به این معنی که در این خوشه، پاسخ‌دهندگان مسیر طولانی‌تری را به‌صورت کندتر طی کردند.

نمودار شاخص دیویس بولدین برای دیتاست احراز هویت نیز محاسبه شده و بهترین k برای دیتاست احراز هویت، ۷ است. آمار توصیفی در جدول ۴ نشان‌داده‌شده است.

جدول ۴ تعداد نمونه‌ها و متوسط ویژگی‌ها در هر خوشه در دیتاست احراز هویت

خوشه	تعداد	درصد	dwelling	flight	trajectory	جعل هویت
۱	۱۲۶	۷	۰/۱۱	۰/۷۳	795.31	۰/۵۳۹۷
۲	۳۴۹	۲۱	۰/۱۲	۰/۸۰	498.34	۰/۴۹۸۶
۳	۴۵۹	۲۷	۰/۱۲	۰/۹۲	383.1	۰/۵۰۵۴
۴	۱۹	۱	۰/۱	۰/۸	1137.59	۰/۵۲۶۳
۵	۴۸۹	۲۹	۰/۱۲	۰/۹۷	283.57	۰/۴۹۰۸
۶	۲۱۱	۱۲	۰/۱۱	۰/۸۰	630.73	۰/۵۰۲۴
۷	۴۹	۳	۰/۱	۰/۸۶	966.76	۰/۴۸۹۸

در این قسمت با استفاده از ویژگی‌های داینامیک ماوس و صفحه‌کلید در دیتاست احراز هویت، از روی دیتاست جعل خوبی، احتمال جعل خوبی را تخمین می‌زنیم. در هر دیتاست دودسته ویژگی فضایی و زمانی وجود دارد که در جدول ۵ نشان داده شده‌اند.

جدول ۵ ویژگی‌های فضایی و زمانی در دیتاست‌ها

	دیتاست احراز هویت	دیتاست جعل خوبی
فضایی	trajectory	auc و mad
زمانی	dwelling و flight	rt، madt، velx و vely

ویژگی‌های بیومتریک ماوس و صفحه‌کلید در دو دیتاست دقیقاً یکسان نیستند. ما به‌صورت تقریبی فرض کردیم ویژگی‌های فضایی به هم شبیه هستند؛ بنابراین به طور مثال چنانچه در یک خوشه، مقدار trajectory پایین باشد، جعل خوبی در آن با خوشه‌ای که در آن مقدار auc پایین است، یکسان است و برعکس چنانچه در یک خوشه، مقدار trajectory بالا باشد، جعل خوبی در آن با خوشه‌ای که در آن مقدار auc بالا است، یکسان می‌باشد. بنابراین خوشه‌ها در دیتاست احراز هویت بر اساس trajectory مرتب شدند و خوشه‌ها در دیتاست جعل خوبی بر اساس auc مرتب شدند. سپس باتوجه‌به درصد تعداد نمونه‌ها در خوشه‌ها، یک مقدار جعل خوبی تحت نام (y1) تخمین زده شده است.

۴- بررسی اثر ویژگی‌های مختلف بیومتریک ماوس و صفحه‌کلید در ترکیب با جعل خوبی

برای یافتن اینکه کدام ویژگی‌های بیومتریک در کنار جعل خوبی، می‌توانند احراز هویت را بهبود دهند آزمایش را با حذف یک یا دو ویژگی تکرار کردیم نتایج در جدول ۶ گزارش شده است

جدول ۶ متوسط و انحراف معیار F در آزمایش‌های مختلف برای بررسی تأثیر ویژگی‌های مختلف

dwell_avg	flight_avg	traj_avg	Y2	متوسط F	انحراف معیار F	
✓	✓	✓		۰/۶۵۰۲	۰/۰۵۸۹۱	
✓	✓	✓	✓	۰/۶۵۰۴	۰/۰۵۸۶	۰/۰۳۳۹
✓	✓			۰/۶۵۰۰	۰/۰۵۹۱	
✓	✓		✓	۰/۶۵۰۰	۰/۰۵۹۲	۰
	✓	✓		۰/۶۵۲۱	۰/۰۴۷۹	
	✓	✓	✓	۰/۶۵۲۳	۰/۰۴۷۶	۰/۰۴۱۷
✓		✓		۰/۶۶۶۲	۰/۰۳۵۱	
✓		✓	✓	۰/۶۶۶۱	۰/۰۳۵۸	–
✓				۰/۶۶۶۳	۰/۰۳۵۱	۰/۰۸۵۴
✓			✓	۰/۶۶۶۶	۰/۰۳۶۱	
	✓			۰/۶۵۱۹	۰/۰۴۸۲	
	✓		✓	۰/۶۵۱۸	۰/۰۴۸۷	–
		✓		۰/۶۶۸۲	۰/۰۰۲۳	
		✓	✓	۰/۶۶۶۸	۰/۰۰۷	–

برای به‌دست‌آوردن اندازه تأثیر از فرمول دلتای گلاس استفاده شد. نتایج نشان داد که ویژگی‌های جعل خوبی در کنار سرعت فشردن کلید (dwelling) بیشترین تاثیر را روی بهبود احراز هویت دارند.

نتیجه‌گیری

احراز هویت[1] یا اصالت‌سنجی فرایندی است که در آن یک فرد باید هویتی را که مدعی آن شده است، ثابت کند. احراز هویت مبتنی بر ریسک یکی از انواع احراز هویت است که بر مبنای میزان ریسکی که از رفتارهای غیرعادی و داده‌های غیرعادی کاربر بدست می‌آورد سعی در احراز هویت کاربر می‌کند هر چه میزان این ریسک بالاتر باشد سیستم از چالش‌های قوی‌تری برای احراز هویت کاربر استفاده می کند. از سوی دیگر، افراد جاعل ممکن است رفتار دستخط صفحه‌کلید و حرکت ماوس را از روی کاربران واقعی تقلید و نمونه برداری کنند. در اینجا مسئله مهمی مطرح می‌شود که آیا کاربر موردنظر در پشت سیستم هست یا فرد جاعل به تقلید رفتار آن می‌پردازد. در این پژوهش هدف بررسی تشخیص افراد جاعل از روی دستخط صفحه‌کلید و حرکات موس انجام شده است. افراد جاعل ممکن است به زمان بیشتری برای پاسخ به سؤالات شریک گفتگو نیاز داشته باشند، یا ممکن است از اصلاحاتی برای تغییر متن خود استفاده کنند تا از تناقض در پاسخ‌هایشان جلوگیری شود. این مسائل باعث ایجاد تفاوت‌هایی در رفتار تایپ می‌شود که می‌توان آن را در ریتم تایپ اندازه‌گیری کرد. می‌توان بادقت بالایی بین چت فردی که در پاسخ‌هایش صادق است و چت فردی که دروغ می‌گوید، تمایز قائل شد.

ازآنجاکه دیتاست آماده‌ای که حاوی ترکیب اطلاعات «جعل خوبی» و «احراز هویت» باشد، در دسترس نبود؛ میزان جعل خوبی از روی خوشه‌های رفتارهای دینامیک ماوس و صفحه‌کلید تخمین زده شد و این تخمین‌ها به‌صورت یک ویژگی «احتمال جعل خوبی» به دیتاست «احراز هویت» اضافه شد. نتایج نشان داد که ترکیب جعل خوبی با ویژگی «سرعت فشردن کلید»، نسبت به سایر ویژگی‌ها، بهبود بیشتری در احراز هویت انجام می‌دهد. بااین‌وجود میزان افزایش معیار F، حداکثر ۰/۱۰ رضایت‌بخش نیست. نیاز است ترکیب‌های بیشتری از دیتاست‌های مختلف مورد بررسی قرار گیرد یا حتی دیتاستی به‌صورت خاص برای انجام پژوهش جمع‌آوری شود.

مراجع

[1] R. Banerjee, S. Feng, J. S. Kang, and Y. Choi, "Keystroke patterns as prosody in digital writings: A case study with deceptive reviews and essays," in Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1469-1473.

[2] M. Monaro et al., "Detecting faking-good response style in personality questionnaires with four choice alternatives," Psychological research, vol. 85, no. 8, pp. 3094-3107, 2021.

[3] M. Szymkowski, P. Milewski, and K. Saeed, "Fingerprint and Keystroke Dynamics Fusion in Multimodal Biometrics System," Advanced Computing and Systems for Security: Volume Eleven, pp. 67-82, 2021.

[4] Y. Li, B. Zhang, Y. Cao, S. Zhao, Y. Gao, and J. Liu, "Study on the BeiHang keystroke dynamics database," in 2011 International Joint Conference on Biometrics (IJCB), 2011, pp. 1-5: IEEE.

[5] C. Mazza et al., "Use of mouse-tracking software to detect faking-good behavior on personality questionnaires: an explorative study," Scientific reports, vol. 10, no. 1, pp. 1-13, 2020.

[1] Authentication