شناسایی جملات حاوی کلمات توهین‌آمیز با استفاده از الگوریتم‌های یادگیری ماشین در سرویس‌های ابری آمازون

امیرعلی خانه‌عنقا[۱]، زهرا موحدی[۲]

[۱] کارشناسی مهندسی کامپیوتر، دانشگاه تهران
amiralikhanehanqa@ut.ac.ir
[۲] استادیار ، گروه مهندسی کامپیوتر، دانشکده مهندسی دانشکدگان فارابی دانشگاه تهران
zmovahedi@ut.ac.ir

چکیده

با گسترش فضای مجازی، نظارت بر این فضا در جهت حفظ ارزش‌های جامعه امری ضروری است. مسئله‌ی استفاده از جملات توهین‌آمیز، زورگویی‌های اینترنتی و استفاده از کلمات مخالف با هنجارهای فرهنگی می‌بایست مورد بررسی و نظارت قرار گرفته و از بروز و نشر آن جلوگیری شود. در این مقاله، راهکاری مبتنی بر پردازش زبان طبیعی ارائه می‌شود تا بتوانیم در زبان فارسی جملات حاوی کلمات توهین‌آمیز را به کمک یادگیری ماشین پردازش کنیم. روش انجام کار به کمک سرویس‌های مختلف ابری آمازون اجرا شده است. نتایج نشان می‌دهد که در مقایسه با روش‌های سنتی، سرویس‌های آمازون موجب تسریع عملیات یادگیری ماشین می‌شوند و توانایی ارائه درصد بالایی از دقت و همچنین پیش‌پردازش و پردازش سریع داده‌ها و استقرار ماشین را دارا می باشند.

کلمات کلیدی: شناسایی کلمات ناهنجار، یاد گیری ماشین، پردازش زبان طبیعی، سرویس‌های ابری آمازون.

۱ مقدمه

امروزه گستردگی فضای مجازی در جهان به قدری است که مکان و زمان را تحت شعاع خود قرار داده و به نوعی هم موجب راحتی و آسایش و هم باعث بروز مشکلاتی از جنس دیگر شده است. مقوله ارتباطات که زمانی رسیدن به نقطه نهایی توسعه آن دغدغه‌ای برای بشر محسوب می‌شد، اکنون با پیشرفت روزافزون علم و تکنولوژی به نقطه‌ای رسیده است که بسیاری از دولت‌ها سعی در مهار افسارگسیختگی این فضا را دارند.

فضای مجازی زمانی در کشور عمومی‌تر شد که متولیان فرهنگی و اجتماعی و مدیران بخش‌های مرتبط با فناوری اطلاعات هیچگونه پیش‌زمینه‌ی لازم به جامعه در خصوص پیامدهای منفی و مخرب فضای مجازی ایجاد نکرده بودند و تا پیش از گستردگی اینترنت و شبکه‌های اجتماعی، کاربران این فضا اکثرا اساتید، نخبگان

و سازمان‌هایی بودند که اینترنت لازمه پژوهش و ارتباطات آنان با سایر نقاط دنیا بود. زمانی که بحث اینترنت به عوام تسری یافت و فراوانی گوشی‌های همراه پیشرفته، فضای مجازی را از حالت اختصاصی خارج کرد بدون اینکه هیچ آموزش و آگاهی در این خصوص به کاربران ارائه شود. نکته‌ای که قابل تاکید می‌باشد این است که تمامی کشورها مدیریت و حاکمیت خاصی بر فضای سایبری اعمال کرده و چارچوبی را برای فعالیت‌ها در این فضا معین نموده‌اند چون آنان نیز به این نتیجه واقف شده‌اند که افسار گسیختگی فضای مجازی در طولانی‌مدت می‌تواند به آسیب‌های جدی اجتماعی دامن زده و بر فرهنگ مردم تاثیر بگذارد و این در حالی است که در صورت کنترل شدن فضای مجازی این بخش قادر خواهد بود جنبه‌های مثبتی را برای کاربران مشخص کرده و رشد و توسعه را برای جامعه رقم بزند.

در فضای مجازی افراد متعددی وجود دارند که صحبت‌ها و کلماتی را بیان می‌کنند که در زندگی واقعی‌شان کمتر از این حرف‌ها از آن‌ها شنیده می‌شود و احساس بی‌بند و باری در فضای مجازی بیشتر نمون پیدا می‌کند؛ محققان به این پدیده مهارگسیختگی (disinhibition) می‌گویند، بدان معنا که افراد کمتر شرم از بیان احساسات خود دارند و از طرفی راحت‌تر شروع به صحبت‌های بی‌ادبانه و پراکندن خشم، نفرت، انتقادهای زننده و حتی تهدید می‌کنند. توانایی در مخفی‌شدن پشت اکانت کاربری، اطمینان از مواجه‌نشدن با افراد در حال مکالمه و یا دور بودن از زمان ملاقات حضوری و بازی انگاشتن فضای مجازی باعث بروز چنین رخدادهایی خواهد شد که نیاز به ایجاد قوانین و بسترسازی جهت جلوگیری از تخریب فرهنگی یک کشور را به امری ضروری مبدل می‌سازد. با توجه به شناخت معضل، بررسی راهکارها جهت جلوگیری از رویداد آن‌ها امری مهم تلقی شده است که در این زمینه یکی از تکنولوژی‌های به‌روز که می‌تواند به حل مسئله یاری رساند، پردازش زبان طبیعی خواهد بود.

منظور از پردازش زبان طبیعی این است که رایانه‌ای داشته باشیم که قادر باشد زبان انسان را تحلیل کند، بفهمد و بتواند زبان طبیعی تولید کند. در نتیجه استفاده از پردازش زبان طبیعی، موجب آماده‌سازی ماشین هوشمند جهت یادگیری و تمییز جملات خوب از بد می‌شود.

در این مقاله، ما دو مدل یادگیری ماشین را در ابتدا برای این مسئله پیاده‌سازی کرده و سپس با کمک سرویس‌های ابری آمازون مدل دیگری را بدست آورده و نتایج آن‌ها را مقایسه می‌کنیم. در انتها مدل به دست آمده را پیاده‌سازی کرده و نتایج آن را با مدل اولیه آمازون مقایسه نموده‌ایم.

در ادامه ساختار مقاله به این صورت است: بخش دوم، کارهای پیشین مورد بررسی قرار گرفته است. در بخش سوم، توصیف روش پیشنهادی آورده شده است. بخش چهارم، پیاده‌سازی روش پیشنهادی و ارزیابی نتایج آن و مقایسه با کارهای پیشین می‌باشد. در بخش پنجم، نتیجه‌گیری و کارهای آینده مورد بحث قرار گرفته است.

۲ مروری بر کارهای گذشته

با توجه به گستردگی موضوع نظارت بر محتوای فضای مجازی ، شاهد مقالات با موضوعات مشابه در خصوص زبان‌های مختلفی هستیم که به ترتیب، زبان آلمانی [۱]، زبان انگلیسی [۲]، زبان ترکی [۳] و زبان اردو [۴] به

جدول ۱: مقایسه پارامترهای استخراج‌شده در حوزه شناسایی کلمات توهین‌آمیز با کمک مدل‌های بررسی‌شده

مقاله مرجع	خلاصه کارهای انجام‌شده	روش استفاده‌شده	پارامتر بررسی‌شده
[۵]	مقاله رویکردی برای شناسایی محتوای توهین‌آمیز به دو زبان انگلیسی و فارسی ارائه شده است.	در این پژوهش، ترکیب BERT با ANN، CNN و RNN بررسی شده است.	F1 macro: 0.863
[۶]	مقاله رویکردی برای شناسایی محتواهای نفرت‌آمیز در زبان انگلیسی و سپس به صورت چندزبانه ارائه شده است.	در این مقاله ابتدا جملات در زبان انگلیسی به کمک مدل Bert پردازش شده‌اند؛ سپس چند زبان دیگر به خصوص زبان فارسی با مدل‌های از پیش پردازش‌شده با مدل Bert بررسی شده‌اند.	F1 macro: 0.878

انجام کارهای مشابه پرداخته‌اند که نشان از اهمیت موضوع دارد. در مقاله‌های یادشده ابتدا به چالشی‌بودن مسئله زبان مهاجمانه در زبان‌های مختلف پرداخته شده و سپس با کمک مدل‌های مختلف به کمک پردازش زبان طبیعی به حل مسئله پرداخته‌اند که نشان از اهمیت بالای این موضوع در زبان‌های مختلف دارد.

در زبان فارسی چالش پیش رو چندان مورد توجه قرار نگرفته و داده‌های عمومی برای آن آماده نشده است؛ با این حال [۵]، یک راهکار دوزبانه ارائه داده است و مسئله را همزمان برای دو زبان فارسی و انگلیسی به کمک مدل Bert و سایر مدل‌ها بررسی کرده است که داده‌ی فارسی آن از ۴۹۸۸ ورودی مختلف تشکیل شده است. طبق پارامترهای بررسی‌شده در مقاله، بهترین متد بیان‌شده، مدل شبکه عصبی کانوولوشن با امتیاز $F1$، ۰٫۸۶۳۰ و $Loss$، ۰٫۴۶۱۷ بیان شده است. همچنین [۶] به این مبحث به صورت چند زبانه پرداخته است که البته یک داده ساختار تشکیل شده از ۶۰۰۰ توئیت فارسی را مورد بررسی قرار داده است، متدولوژی این تحقیق نیز بر روی مدل Bert و شبکه‌های عصبی کانوولوشن پایه‌گذاری شده است و چندین مدل مختلف با یکدیگر به رقابت گذاشته شده‌اند. جدول ۱ خلاصه کارهای پیشین در زبان فارسی را نشان می‌دهد.

۳ روش‌های پیشنهادی

در روش‌های پیشنهادی، ابتدا دو مدل k-نزدیک‌ترین همسایه و مدل شبکه عصبی با توجه به پژوهش‌های انجام‌گرفته بررسی شده‌اند؛ سپس به کمک سرویس AutoML، مدل پیشنهادی سرویس آمازون را بررسی کرده و مدل به دست آمده را شبیه‌سازی کرده تا بتوانیم نتایج پیشین را بررسی و مقایسه کنیم.

۱.۳ مدل k-نزدیک‌ترین همسایگی

طبق [۷]، الگوریتم k-نزدیک‌ترین همسایگی (KNN) برای مسائل طبقه‌بندی و رگرسیون قابل استفاده است. اگرچه در اغلب مواقع از آن برای مسائل طبقه‌بندی استفاده می‌شود. برای ارزیابی هر روشی به طور کلی به سه جنبه مهم آن توجه می‌شود: ۱) سهولت تفسیر خروجی‌ها، ۲) زمان محاسبه و ۳) قدرت پیش‌بینی الگوریتم KNN در سه مرحله لیبل داده جدید را تخمین می‌زند:

۱. در مرحله اول KNN، فاصله نمونه تست را با تمامی نمونه‌های آموزش محاسبه می‌کند.

۲. سپس در مرحله‌ی دوم بر اساس فاصله بدست آمده، k تا نزدیک‌ترین همسایه از داده‌ی آموزش به داده‌ی تست را پیدا می‌کند.

۳. در مرحله‌ی آخر رای‌گیری انجام می‌دهد تا متوجه شود که از بین این k تا نزدیکترین همسایه کدام کلاس بیشترین همسایه از داده‌ی آموزش به نمونه تست دارد تا تصمیم بگیرد که نمونه جدید به آن کلاس تعلق دارد.

جهت آماده‌سازی داده‌ها، در ابتدا می‌بایست با استفاده از رمزگذار لیبل‌ها، داده‌های جمع‌آوری‌شده جهت آموزش و تست به دیتافریم قابل استفاده در کدنویسی تبدیل شوند. سپس قسمت‌های مختلف داده‌ی تست که دارای لیبل خوب یا بد بودن جملات است از هم جدا شده و در دیتافریم‌های مختلف ذخیره می‌گردند و خروجی‌ها برای استفاده در مدل، تبدیل به داده‌ی ساختار لیست می‌شوند. بخش بعدی انتقال داده به حالت استاندارد و تبدیل آن‌ها به ویژگی‌های قابل استفاده توسط ماشین است؛ بدین معنی که داده‌ها باید یک ارزش اولیه به خود گرفته و عادی‌سازی انجام شود. سپس یک لیست از داده‌ها و مقدار مربوطه ساخته می‌شود تا بتوان پردازش را بر روی آن‌ها انجام داد. در بخش بعدی، نوبت به انجام عملیات یادگیری ماشین به کمک مدل توضیح‌داده‌شده رسیده است. ابتدا مشخص می‌کنیم چند درصد از داده تست به عنوان آموزش و چند درصد برای ارزیابی قرار بگیرد؛ سپس به کمک مدل ازپیش‌آماده k-نزدیک‌ترین همسایه را فراخوانی کرده و داده‌ی آماده‌شده به ورودی آن داده شده است. این مدل نیاز دارد بداند که فاصله بین ویژگی‌ها را در چه ابعادی بررسی کند. پس از اتمام پردازش یادگیری ماشین، همانگونه که داده‌های آموزشی انتقال یافتند، داده‌های تست نیز آماده و برای پیش‌بینی به مدل داده می‌شوند و خروجی مورد نظر جهت ارزیابی ذخیره خواهد شد.

۲.۳ مدل شبکه عصبی

شبکه‌های عصبی مجموعه‌ای از نورون‌ها هستند که از الگوریتم‌های منحصر به فردی پیروی می‌کنند [۸]. این مجموعه که از مغز انسان الگوبرداری و الهام گرفته شده است، با هدف شناسایی الگوها طراحی می‌شود و مورد استفاده قرار می‌گیرد. به طور کلی می‌توان گفت که شبکه عصبی شامل الگوریتم‌هایی‌ست برای یادگیریِ ماشین، که منجر به طبقه‌بندی کردن داده‌های ورودی و ارائه خروجی مطلوب می‌گردد. به همین دلیل است که می‌توان شبکه‌های عصبی را به عنوان جزئی از فرایند یادگیری ماشین در نظر گرفت.

برای شناخت ساختار شبکه باید دانست، این شبکه از دو قسمت اصلی تشکیل شده است. قسمت اول لایه‌ای با عنوان Embedding می‌باشد که به کمک آن بُرداری از اعداد متناسب با هر کلمه ذخیره می‌شوند. به معنای دیگر در این لایه برای هر کلمه یک بردار از اعداد نسبت داده می‌شود که اعداد داخل این بردار در حین آموزش شبکه تغییر می‌کنند. تعدادی از این لایه‌های Embedding ازقبل‌آموزش‌دیده موجود است که می‌توان از آن‌ها استفاده کرد و دیگر نیازی به آموزش شبکه نباشد؛ ولی به دلیل این که داده‌های موجود، فارسی می‌باشند نمی‌توانیم از Embeddingهای آماده استفاده کنیم و نیاز است که شبکه را آموزش بدهیم. قسمت دوم شبکه را لایه‌های LSTM تشکیل می‌دهند. از این لایه‌ها اغلب برای داده‌های سری زمانی استفاده می‌کنند که در زمینه‌ی داده‌کاوی و پروژه‌های مشابه نیز کاربرد دارند. پس از آموزش شبکه موجود برای تست شبکه، نمونه داده‌هایی را وارد کرده و خروجی بررسی خواهد شد. نتیجه تست شبکه آموزش‌دیده را بر روی داده‌های تست می‌توان مشاهده کرد.

۴ پیاده‌سازی و نتایج

در این بخش، دو حالت پیاده‌سازی را مورد بررسی قرار می‌دهیم. حالت اول مدل k-نزدیک‌ترین همسایه و شبکه عصبی جهت تشخیص کلمات توهین‌آمیز می‌باشد که به صورت لوکال پیاده‌سازی و اجرا شده است و حالت دوم پیاده‌سازی و اجرا توسط سرویس‌های آمازون است.

۱.۴ پیاده‌سازی الگوریتم‌های پیشنهادی

داده‌های جمع آوری شده تعداد ۶۰۰ جمله هستند که قسمتی از آن‌ها توهین‌آمیز و قسمتی دیگر جملات مودبانه می‌باشند. البته دیتاست موجود به سه قسمت سالم، ناسالم و بسیار ناسالم تقسیم شده است ولی برای راحتی کار دو دسته‌ی آخر در گروه ناسالم قرار گرفته شده‌اند. تمامی داده‌ها (جملات) در فایل اکسل قرار می‌گیرند. در فایل اکسل علاوه بر جملات در ستون جداگانه‌ای با عنوان "Label" مقدار صفر برای داده‌های ناسالم و یا یک برای جملات سالم اختصاص داده شده است؛ در واقع با این کار داده‌های هدف خود را ایجاد می‌کنیم که برای آموزش شبکه مورد نیاز می‌باشند. در ادامه داده‌ها خوانده و در یک جدول ذخیره شده‌اند تا برای ورود به مدل‌های ارائه‌شده آماده باشند. پردازش‌های پیشنهادشده در بخش ۳، به کمک پردازنده Core i7-10750H با قدرت پردازشی 2.60 GHz انجام شد؛ همچنین از یک حافظه Ram، 16 GB استفاده شده است. پیاده‌سازی انجام‌شده، به کمک زبان برنامه‌نویسی پایتون، و به کمک کتابخانه‌های Pandas [۹] که در جهت خواندن فایل‌ها و Scikit Learn [۱۰] و Keras [۱۱] برای پیاده‌سازی مدل‌های مختلف استفاده شده‌اند. با توجه به حجم داده‌ی سرعت پردازش‌ها به مرور کاهش می‌یابد، به هر میزان داده افزایش یابد با در نظر گرفتن محدودیت‌های سخت افزاری آموزش داده‌ها زمان بیشتری را خواهد طلبید. در ادامه استفاده از سرویس ابری آمازون مورد بحث قرار گرفته است.

۲.۴ استفاده از سرویس‌های یادگیری ماشین آمازون با داده‌های موجود

Amazon SageMaker Studio یکی از معروف‌ترین سرویس‌های آمازون کلود برای آموزش و استقرار مدل‌های مختلف یادگیری ماشین است. Autopilot یکی از ابزارهای سرویس SageMaker است که وظایف کلیدی فرایند یادگیری ماشین را خودکار می‌کند؛ یعنی داده‌ها را بررسی کرده، الگوریتم‌های مربوط به نوع مسئله را انتخاب و داده‌ها را برای آموزش و تنظیم مدل آماده می‌کند [۱۲].

Best model	Accuracy (Objective)	PrecisionMacro	BalancedAccuracy	F1macro	RecallMacro	Algorithm
OffensivePersianLanguageTest2bJL-083-c4e9819e	0.962	0.968	0.933	0.947	0.933	XGBoost

1 row selected

Model name	Objective: Accuracy	F1macro	Status
OffensivePersianLanguageTest2bJL-08... Best model	0.962	0.947	Completed
OffensivePersianLanguageTest2bJL-084-b5e2ac8a	0.962	0.945	Completed
OffensivePersianLanguageTest2bJL-063-270278b5	0.961	0.944	Completed
OffensivePersianLanguageTest2bJL-081-3ecda2f2	0.961	0.944	Completed
OffensivePersianLanguageTest2bJL-096-5172b861	0.961	0.946	Completed
OffensivePersianLanguageTest2bJL-065-330fd40f	0.959	0.943	Completed
OffensivePersianLanguageTest2bJL-094-3058535b	0.959	0.944	Completed
OffensivePersianLanguageTest2bJL-059-4afd9843	0.959	0.941	Completed
OffensivePersianLanguageTest2bJL-091-6e523b67	0.959	0.941	Completed

شکل ۱: بخشی از مدل‌های مختلف بررسی‌شده به صورت اتوماتیک و انتخاب مدل XGBoost به عنوان بهترین مدل و درصد دقت و F1

در مرحله اول داده‌ها باید در Amazon S3 قرار بگیرند تا به آن‌ها از سمت Amazon SageMaker Studio دسترسی داده شود. در مرحله بعد، از بخش فایل گزینه Experiment را انتخاب می‌کنیم تا به محیط ساخت یک پایپلاین Autopilot هدایت شویم. در ادامه مشخص می‌کنیم که کدام ستون هدف ما خواهد بود تا استودیو بتواند داده را تجزیه تحلیل کرده و بر اساس آن یک مدل پیشنهاد دهد. در انتها نیاز است تعیین گردد پروژه از چه نوعی است و چه نوع پردازشی براساس داده‌ها بهتر است که البته دارای نوع خودکار نیز می‌باشد. به ازای هرکدام از این مراحل داده‌های مخصوص به آن را آمازون استخراج کرده و در یک بخش مخصوص به خروجی آن قرار می‌دهد و برای بررسی بهتر قابل دسترسی خواهد بود. در ادامه سیستم Autopilot، شروع به مهندسی ویژگی می‌کند. مهندسی ویژگی یا استخراج ویژگی، فرایند استفاده از دانش دامنه برای استخراج ویژگی‌ها از داده‌های خام است. انگیزه‌ی استفاده از این ویژگی‌های اضافی برای بهبود کیفیت نتایج حاصل از فرایند یادگیری ماشین در مقایسه با ارائه تنها داده خام به فرایند یادگیری ماشین می‌باشد. در این راه آمازون داده‌های با ساختار بهتر برای استفاده در مدل‌ها ایجاد می‌کند.

در بخش اصلی پردازش سرویس، مدل‌های مختلف موجود با داده‌های پردازش‌شده تحلیل می‌گردند و دقت آن‌ها ارزیابی شده و در نهایت بهترین مدل توسط سرویس معرفی می‌گردد. همچنین در این مرحله هر بار مدل‌ها تنظیم می‌شود، تنظیم معمولاً یک فرآیند آزمون و خطا است که توسط آن برخی از فراپارامترها تغییر داده می‌شوند (مثلاً تعداد درخت‌ها در یک الگوریتم مبتنی بر درخت یا مقدار آلفا در یک الگوریتم خطی)، دوباره الگوریتم روی داده‌ها اجرا می‌شود، سپس عملکرد آن در مجموعه اعتبارسنجی مقایسه می‌شود تا

مشخص شود کدام مجموعه از فراپارامترها دقیق‌ترین مدل را به دست می‌آورند. با این حساب، دقت مدل تا جای امکان توسط خود سامانه Autopilot انجام می‌گیرد.

همان‌طور که در شکل ۱ مشخص است، بهترین مدل موجود از نظر آمازون مدل XGBoost است. الگوریتم Extreme Gradient Boosting از دسته الگوریتم‌های گرادیان تقویتی بوده که عملکرد بسیار خوبی در دسته‌بندی، رگرسیون و رتبه‌بندی داشته و به دلیل پیش‌بینی دقیق، سرعت زیاد و پشتیبانی از اجرای چندمنظوره و توزیع‌شده‌ی آن، در مسائل دسته‌بندی بسیار محبوب است. XGBoost به طور خاص، این الگوریتم را برای تقویت درخت تصمیم با یک عبارت تنظیم سفارشی اضافی در تابع هدف پیاده‌سازی می‌کند.

۳.۴ پیاده‌سازی سرویس مدل XGBoost

با توجه به اطلاعات بدست‌آمده، الگوریتم XGBoost [۱۳] را همانند مدل‌های پیشنهادی در بخش ۳، پیاده‌سازی کرده‌ایم تا علاوه بر مقایسه با کارهای پیشین در این مقاله و همچنین مقایسه با کارهای پیشین در فصل دو بتوانیم این اطلاعات را با یک پیاده‌سازی توسط عامل انسانی نیز مقایسه کنیم. پیاده‌سازی این مدل به کمک کتابخانه XGBoost انجام شد و مراحل آن همچون مدل KNN می‌باشد.

با وجود دانش بر استفاده از مدل XGBoost، بدون تنظیم دقیق مدل بر اساس داده و پیش‌پردازش دقیق به درصدی پایین‌تر از خروجی مدل آمازون و با مقادیری نزدیک به مدل‌های قبلی خواهیم رسید که نشان از تاثیرگذاری پردازش‌های آمازون بر روی متن و همچنین شخصی‌سازی مدل برای داده را دارد.

با این وجود برای پردازش داده‌های خاص و پر اهمیت نیاز به یک متخصص علم داده برای فهم و ایجاد تغییرات مورد نیاز است. همچنین با توجه به استفاده آمازون از کتابخانه‌های اختصاصی خود متخصص در صورت نیاز به انجام پروژه‌های خاص‌تر و یا شخصی‌سازی‌شده‌تر نیاز به آشنایی با نحوه کار با این کتابخانه‌ها خواهد داشت و یادگیری آن‌ها می‌تواند زمان یادگیری برای انجام کار را بیش از پیش کند. همچنین در صورت نیاز به استقرار محصول، معمول‌ترین راه استفاده از مدل‌ها استفاده از خود آمازون است و در صورت نیاز به استقرار به صورت محلی چالش‌هایی وجود دارد. در نهایت به نظر می‌آید وجود محصولات نوین در سرویس‌های ابری می‌تواند کمک شایانی به رشد علوم مختلف کند.

۴.۴ ارزیابی نتایج

در این پژوهش تلاش شد تا به بررسی کنترل محتوای متنی فضای سایبری با گرایش پردازش زبان پرداخته شود تا مزایای روش‌های جدید نسبت به روش‌های معمول نشان داده شود. هدف از پیدایش سرویس‌هایی همچون Amazon Autopilot کمک به صنایعی است که علاقه‌مند به رشد خود به کمک هوش مصنوعی را دارند. از این طریق این علم به‌روز نیز فرصت‌های بیشتری برای پیشرفت پیدا خواهد کرد. همچنین کاربران نهایی نیز می‌توانند از پیشنهادات و همچنین سهولتی که توسط صاحبان صنایع برایشان ایجاد کرده‌اند، استفاده کرد.

بایستی اشاره کرد همچنان نیز دانشمندان علم داده و محققان عرصه هوش مصنوعی گام‌های متعددی

Metrics table

Metric Name	Value	Standard Deviation
accuracy	0.9625	0.0036570983477315945
weighted_recall	0.9625	0.003657098347731561
weighted_precision	0.9641346500721502	0.0034308255123618463
weighted_f0_5	0.9627209249918511	0.0036665997672579364
weighted_f1	0.9616973988367469	0.003823846455268386
weighted_f2	0.9618344359589286	0.003778078644836735
accuracy_best_constant_classifier	0.496875	0.005375001455356815
weighted_recall_best_constant_classifier	0.496875	0.005375001455356815
weighted_precision_best_constant_classifier	0.246884765625	0.005316383722923077
weighted_f0_5_best_constant_classifier	0.2745070361362057	0.005588829785636524
weighted_f1_best_constant_classifier	0.3298669102296451	0.005937197673386191
weighted_f2_best_constant_classifier	0.4131962604602511	0.005961186581812905

شکل ۲: اطلاعات آماری قابل برداشت از سرویس Amazon Autopilot

جهت اعتلای این علوم نوین در پیش خواهند داشت و بیشتر این ابزار جهت تسهیل راه آن‌ها در برداشتن این گام‌ها خواهد بود. جدای از زمان مورد نیاز برای جمع‌آوری داده‌های قابل استفاده برای یک پروژه‌ی یادگیری ماشین، انجام مراحل پیش‌پردازش، پردازش و استقرار یک ماشین آموزش‌دیده در آمازون در برابر طی‌کردن مسیر به طور معمول بسیار کمتر بود؛ به نحوی که ششصد داده‌ی آماده‌شده توسط سرویس‌های این مجموعه کمتر از دو ساعت به طول انجامید؛ در حالی که انجام موارد مستلزم دو مدل قبلی و پردازش به کمک آن‌ها زمان بسیار بیشتری از نظر پیدا کردن الگوی مناسب و همچنین پیش‌پردازش نیاز داشت و تمامی کار در حدود یک ماه به طول انجامید که البته در صورت وجود یک متخصص ممکن بود بسیار کاهش پیدا کند. علاوه بر این، نیاز صنایع برای به‌کارگیری متخصصان حرفه‌ای با توجه به ابزارهای موجود در آمازون کاهش پیدا می‌کند.

شایان ذکر است از لحاظ کارآمدی مدل‌های آموزش‌دیده KNN و شبکه عمیق نسبت به مدل‌های آماده‌شده و تنظیم‌شده توسط آمازون دقت کمتری را ارائه داده‌اند.

۵.۴ ارزیابی نتایج

همان‌طور که در جدول ۲ مشخص شده است، دقت مدل پردازش‌شده توسط آمازون از مدل‌های آماده‌شده‌ی قبلی بالاتر بوده و نشان‌دهنده‌ی دقت بالای این سرویس است. با این وجود آمازون قابلیت این را دارد که کاربر مدل را تحلیل کرده و در صورت نیاز تغییراتی در آن ایجاد کند.

۵ نتیجه‌گیری

در این مقاله به بررسی شناسایی کلمات توهین‌آمیز در خطوط متنی فارسی زبان با هدف نظارت بر سلامت اخلاقی فضای مجازی پرداختیم. در راستای این هدف، دو مدل یادگیری ماشین نزدیک-k همسایه و مدل شبکه عصبی پیشنهاد شد. سپس با کمک سرویس Amazon SageMaker و ابزار Autopilot این سرویس بر روی دیتاست موجود، مدل XGBoost به عنوان مدل برتر با دقت بهتر به دست آمد. در نهایت جهت اعتبارسنجی نتیجه‌ی سرویس Autopilot آمازون، این مدل نیز پیاده‌سازی و با دو مدل قبلی مقایسه گردید.

جدول ۲: مقایسه پارامترهای استخراج‌شده در حوزه شناسایی کلمات توهین‌آمیز با کمک مدل‌های بررسی‌شده

مدل	دقت	F1 macro
KNN	0.875	0.594
شبکه عصبی	0.835	0.292
Amazon Autopilot XGBoost	0.962	0.933
XGBoost	0.860	0.561

نتایج به‌دست آمده حاکی از برتری این مدل و در نتیجه اعتماد به سرویس Autopilot آمازون است.

مراجع

[1] Samantha Kent. (2018). German hate speech detection on Twitter, In Proceedings of the 14th Conference on Natural Language Processing (KONVENS'18),120–124.

[2] Georgios Pitsilis, Heri Ramampiaro, and Helge Langseth, (2018). Effective hate-speech detection in Twitter data using recurrent neural networks, Appl, Intell, 48, 4730–4742.

[3] Selma Ayse Ozel, Esra Sarac, Seyran Akdemir, and Hulya Aksu, (2017). Detection of cyberbullying on social media messages in Turkish, In Proceedings of the International Conference on Computer Science and Engineering (UBMK'17), 366–370.

[4] Raza Mustafa, M. Saqib Nawaz, Javed Ferzund, M. Ikram Ullah Lali, Basit Shahzad, and Philippe Fournier-Viger, (2017). Early detection of controversial Urdu speeches from social media. Data Sci, Pattern Recog, 1, 26–42.

[5] Marzieh Mozafari. (2021). Hate speech and offensive language detection using transfer learning approaches. Document and Text Processing, Institut Polytechnique de Paris.

[6] Alavi, P., Nikvand, P., & Shamsfard, M. (2021). Offensive Language Detection with BERT-based models, By Customizing Attention Probabilities. ArXiv, abs/2110.05133.

[7] Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K. (2003). KNN Model-Based Approach in Classification. In: Meersman, R., Tari, Z., Schmidt, D.C. (eds), On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE. OTM 2003. Lecture Notes in Computer Science, vol 2888. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-39964-3_62.

[8] Alshemali, B., & Kalita, J. (2019). Improving the reliability of deep neural networks in NLP: A review. Knowledge-Based Systems. https://doi.org/10.1016/j.knosys.2019.105210.

[9] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, .2825-2830

[10] McKinney, W. (2011). Pandas: a foundational Python library for data analysis and statistics. Python for high performance and scientific computing, 14(9), .1-9

[11] Ketkar, N. (2017). Introduction to keras. In Deep learning with Python (pp. 97-111). Apress, Berkeley, CA.

[12] Antje Barth, Shelbee Eigenbrode, Sireesha Muppala, Chris Fregly, Analyze Datasets and Train ML Models using AutoML, Coursera. https://www.coursera.org/lecture/automl-datasets-ml-models/specialization-overview-eqDJv.

[13] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K. (2015). XGBoost: extreme gradient boosting. R package version 0.4-2, 1(4), .1-4