

اتوماسیون استخراج شبهات در شبکه اجتماعی توییتر با هوش مصنوعی

سیناصیفوری

این مقاله با هدف استخراج شبهات از شبکه اجتماعی توییتر، نقشه راهی برای چگونگی انجام این کار با استفاده از علوم کامپیوتر بخصوص هوش مصنوعی ارائه داده است. این اتفاق دسترسی به شبهات را بسیار ساده تر کرده و در زمان کارشناسان از حیث پیدا کردن، دسته بندی کردن شبهات و پاسخگویی به آنها صرفه جویی می کند. با استفاده از این سیستم امکان تحلیل آماری بر روی شبهات و زمینه هایی که بیشترین بحث را برای جامعه به وجود آورده اند به همراه پاسخ هایی که کارشناسان به این سوالات داده اند زمینه را برای پیدا کردن بهترین پاسخ از نظر پذیرفته شدن اجتماعی فراهم نموده است. در این فرآیند تکنولوژی های هوش مصنوعی، اتوماسیون سازی و پردازش متن مورد استفاده قرار گرفته اند و بخشی های مهمی از کار یک کارشناس پاسخگو به شبهات را ساده تر نموده اند.

واژگان کلیدی: شبهات، توییتر، هوش مصنوعی، شبکه اجتماعی، یادگیری ماشین

۱ مقدمه

یکی از طبیعی ترین واکنش های انسانی نسبت به هر آموخته جدید که یکی از مراحل درک بهتر مفهوم مسائل نیز می باشد، طرح پرسش ها و شبهات است. دین نیز به عنوان یک آموزه که همه ابعاد زندگی انسان را تحت تاثیر قرار می دهد از این مورد مستثنا نیست و پرسش ها و شبهاتی به همراه دارد که منفی یا مثبت بودن آن به نیت مطرح کنند آن شبهه بستگی دارد. [۹] در دهه اخیر که شاهد گسترش چشمگیر نقش شبکه های اجتماعی در زندگی انسان هستیم، به طوری که شبکه های اجتماعی مختلفی با تمرکز بر بخش خاصی از نیازهای انسانی فعالیت های کاربران را تا حدود زیادی هدفمند کرده اند. شبکه اجتماعی توییتر از آغاز شروع به کار تا به امروز با دنبال کردن هدف گردش آزاد اطلاعات و اخبار توانسته به بستری برای انواع پیام های سیاسی، اقتصادی و اجتماعی شود. [۸]

امروزه جایگاه توییتر به گونه ای است که بسیاری از سیاستمداران بزرگ دنیا اعم از شرقی و غربی به جای مکاتبات رسمی بسیاری از پیام های خود را از طریق این رسانه به گوش جهانیان می رسانند، از جمله مزایای انکارناپذیر این بستر آزاد بودن و امکان نقد و به چالش کشیدن همه اشخاص، حتی ادیان و مکاتب فلسفی می باشد که پیامد مخفی بودن هویت کاربران می باشد. [۴] این سوال که آیا با استفاده از توییتر قادر به استخراج شبهات دینی کاربران ایرانی و یا حتی غیر ایرانی هستیم یا خیر؟ و کم و کیف این موضوع سوال زمینه ای و کلیدی این پژوهش می باشد که به پاسخ آن خواهیم پرداخت.

علت اصلی که تمرکز استخراج شبهات بر توییتر قرار گرفته است، گردش آزاد و خارج از کنترل اطلاعات در این رسانه است که در صورت نادیده گرفته شدن موجب آسیب های جبران ناپذیری خواهد شد. [۷] با توجه به وجود موانعی برای حضور در این شبکه اجتماعی، مطابق فتوای مقام معظم رهبری حضور در این شبکه اجتماعی در صورتی که موجب گناه و یا تقویت دشمنان اسلام نشود منعی ندارد و از آنجا که نیت حضور و استفاده از این شبکه اجتماعی جلوگیری از انحرافات فکری و خنثی سازی شبهاتی که بطور غرض ورزانه منتشر می شوند است، از نظر شرعی منعی برای استفاده از اطلاعات این شبکه اجتماعی وجود ندارد.

نقش هوش مصنوعی در مسیر استخراج شبهات از توییتر مربوط به بخش هایی می شود که باید کامپیوتر تشخیص دهد که آیا یک توییت برگزیده حاوی شبهه است و یا خیر. برای مثال یک توییت که در آن کلمه قرآن وجود دارد باید مورد بررسی قرار گیرد که آیا در وصف قرآن است و یا یک شبهه در مورد قرآن کریم، دسته بندی های دیگری نیز وجود دارد و برای مثال یک توییت می تواند مسئله ای را به یکی از آیات قرآن کریم ارجاع داده باشد و یا موارد دیگری که باید توسط هوش مصنوعی تشخیص داده شده و در صورت شبهه بودن به کارشناسان مربوطه برای پاسخگویی گزارش داده شود. نکته حائز اهمیت دیگر که توسط هوش مصنوعی انجام می شود، دسته بندی شبهات است، یعنی شبهات بسته به شرایط و نیاز به چند دسته اصلی نظیر قرآن کریم، اهل بیت و قرآن کریم تقسیم می شوند که برای پاسخ داده شدن به کارشناس تخصصی هر دسته ارجاع داده شود.

کارشناسان مربوطه می‌رساند. در این روش زمانی جهت پیدا کردن شباهت در توییت‌ها توسط کارشناس محترم صرف نمی‌شود و در نتیجه زمان بیشتری برای پاسخگویی به پرسشها و شباهت دارد.

۳.۲ قالب پژوهش

برای انجام این پژوهش متناسب با موضوع مطرح شده با استفاده از روش پیاده سازی نرم‌افزاری و ثبت وقایع و مراحل با توجه به استاندارد برنامه نویسی پروژه‌های داده کاوی، یعنی CRISP انجام شد و داده‌های ثبت شده با استفاده از مدل‌هایی که در این فرآیند توسعه داده شدند مورد پردازش‌های لازم قرار گرفته و نتایج ثبت گردید. برای استخراج داده از توییت‌ها اولاً باید یک حساب کاربری در توییت ایجاد کرد و سپس ایمیل و شماره همراه را تایید کرد، شیوه تایید ایمیل و شماره همراه بدین صورت است که پس از وارد کردن مقادیر، کدی به شماره همراه شما ارسال می‌شود که باید در توییت وارد کنید تا شماره شما تایید شود و برای ایمیل نیز لینکی به ایمیل شما ارسال می‌گردد و پس از کلیک کردن بر روی آن لینک ایمیل تایید می‌گردد. برای ارتقا نوع حساب کاربری به توسعه دهنده اولاً باید وارد آدرس اینترنتی developer.twitter.com شده و درخواست خود را ثبت نمایید. مهم‌ترین بخش در ثبت درخواست انگیزه نامه می‌باشد که باید به زبان انگلیسی و در مورد هدفی که از استفاده داده‌ها دارید مواردی را بیان کنید. درخواست طی ۲ الی ۴ روز مورد بررسی قرار گرفته و نتیجه به ایمیل حساب کاربری ارسال می‌گردد. مطابق آزمایش‌ها اگر حساب کاربری شما با پیش شماره ایران باشد به دلایل نامعلومی از سوی توییت درخواست شما رد خواهد شد، به همین دلیل برای استفاده از این امکان باید از شماره کشورهایی که تحریم نیستند استفاده نمایید، این شماره‌ها در قالب سرویس‌های شماره مجازی بر بستر اینترنت ارائه شده‌اند.

سپس داده‌های استخراج شده با استفاده از الگوریتم‌های هوش مصنوعی مورد پردازش قرار داده شدند. این مسیر از استخراج این داده‌ها شروع و با ثبت پاسخ‌ها توسط کارشناسان به طور موقت پایان می‌یابد. علت پایان موقتی بازخوردهای ثانویه و سوالات جانبی پیرامون پاسخ مطرح شده می‌باشد که کاربران مطرح کننده شباهت و سوالات از کارشناس مربوطه در قالب کامنت می‌پرسند.

۳.۳ داده کاوی (Data Mining)

امروزه پایگاه‌های داده پیشرفته انواع مختلفی از داده‌ها در خود جای می‌دهند؛ این گستردگی انواع داده‌ها باعث شده روش‌های آماری و مدیریتی ساده برای تحلیل این داده‌ها جوابگو نباشند و استخراج دانش از حجم‌های زیاد داده تبدیل به چالش بزرگی شود.

از دلایل اصلی که استخراج شباهت مطرح شده در توییت‌ها مورد اهمیت قرار دارد، بدیع بودن این نوع از استفاده اطلاعات این شبکه اجتماعی است. یعنی پروژه‌ها و پژوهش‌هایی که در زمینه این شبکه اجتماعی انجام شده‌اند اغلب جنبه تجاری داشته و برای پشتیبانی و پاسخگویی به سوالات مشتریان استفاده شده‌اند. استخراج اطلاعات و ایجاد گزارش از خروجی به دست آمده از این حیث اهمیت دارد که شبکه‌های اجتماعی آینه تمام قدی از احساسات، نیازها و واکنش‌های جامعه بصورت لحظه‌ای نسبت به وقایع و اخبار تزریق شده به جامعه می‌باشد. با توسعه مدل‌های هوش مصنوعی که داده‌های موجود در توییت را تبدیل به اطلاعات قابل فهم و خلاصه سازی شده کنند، بهترین مسیر برای جلب رضایت و افزایش آگاهی کاربران جامعه هدف برای ما مشخص می‌شود. [۱۱]

به بیان دیگر استخراج شباهت از توییت‌ها اولین و ساده‌ترین دستاورد استفاده از پیام‌های توییت در حوزه شباهت دینی می‌باشد، دیگر دستاوردهای این روش مواردی نظیر به دست آمدن نتایج آماری در زمینه بیشترین و کمترین مسائلی که برای جامعه مورد سوال بوده است به همراه پاسخ آن می‌باشد، همچنین از نتایج آماری مشخص می‌شود که افرادی که پیام‌ها و یا شباهت را انتشار می‌دهند بر روی چه موضوعاتی فعالیت می‌کنند، چرا که پس از استخراج، شباهت توسط سیستم طراحی شده به کارشناس ارجاع داده شده و پس از ثبت شدن پاسخ بصورت کامنت بر روی توییت حاوی شبهه از طریق تحلیل تعداد لایک و کامنت‌ها و تحلیل احساس کامنت‌های بازخوردی توسط هوش مصنوعی قانع کننده‌ترین پاسخ و شیوه پاسخ دهی به دست می‌آید. سپس اطلاعات به دست آمده بصورت‌های مختلفی نظیر خوراکی‌های فرهنگی و فرامتن‌های حساب شده، به جهت پیشگیری و یا رفع شبهه ایجاد شده قابل عرضه به جامعه می‌باشند. [۱۳]

۳ روش انجام پژوهش

۳.۱ مقدمه

با گسترش شبکه‌های اجتماعی و فناوری‌های ارتباطی، روش‌هایی که برای اطلاع‌رسانی و رفع شباهت استفاده می‌شده نیز نیز دستخوش تغییراتی شده است. امروزه روش‌های مختلفی نظیر تماس با دفتر مراجع تقلید، پرسش سوال در فرم‌ها و پرتابل‌های اینترنتی از روش‌های مرسوم هستند، اما در این پژوهش قصد آشنایی با روش دیگری را داریم که به طور سازماندهی شده و اتوماسیون شده و سیستماتیک شباهت افراد در شبکه اجتماعی توییت را به گوش

		مقدار پیشبینی شده	
		مثبت	منفی
مقدار واقعی	مثبت	TP	FN
	منفی	FP	TN

تصویر ۱: ترسیم ماتریس اغتشاش دو حالته

پس از ایجاد ماتریس اغتشاش برای پیشبینی‌ها و نتایج صحیح با استفاده از فرمول Accuracy دقت مدل ما به دست می‌آید. برای محاسبه میزان دقت مطابق نامگذاری‌های صورت گرفته در جدول باید از فرمول زیر استفاده کرد.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

و در نهایت پس از انجام این مراحل در صورتی که میزان دقت مورد قبول بود نتایج به دست آمده ثبت و تحلیل و نتیجه‌گیری‌های لازم از روی آن‌ها انجام می‌شود.

۳. پیشبینی (Prediction): در این مرحله مدل آموزش دیده با توجه به داده‌های پیشین و براساس فرمول‌های ریاضیاتی و آماری اقدام به پیشبینی برای داده‌های جدید می‌کند. صحت این پیشبینی‌ها بستگی به دقت مدل و نتایج ارزیابی‌های گام ۲ دارد.

۳.۴ پیش پردازش

در اولین مرحله پژوهش که احتیاج به داده کاوی دارد احتیاج به متنی یکدست و آری از هرگونه کلمات اضافی و شکست‌های متن برای صفحه بندی داریم. در اولین مرحله از پیش پردازش احتیاج به یکدست سازی متن و حذف ایست‌واژه‌ها (Stopwords) که بیشتر آن‌ها همان کلمات دستوری هستند و برخی دیگر نیز علائمی هستند که در زبان‌ها استفاده شده‌اند. و برای مرحله دوم استفاده از هوش مصنوعی علاوه بر موارد بخش اول احتیاج به برچسب‌گذاری داده‌ها با توجه به موضوع تحقیق بود، به علت گسترده شدن کلید واژه‌ها امکان برداشت دسته‌بندی صرفاً بر اساس آن‌ها کلیدواژه‌ها ممکن نبود.

پیش پردازش مربوط به هر بخش از این فرآیند نیازهای نسبتاً مشابهی دارد، به همین علت مرحله پیش پردازش پس از آزمایش در هر بخش و اعلام نتایج در صورت نیاز مورد تغییر قرار می‌گرفت، در صورتی که در هر بخش نسبت به پیش پردازش اصلی تغییری صورت گرفته باشد در همین پژوهش و بخش مربوطه بیان خواهد گردید و موارد تغییر را به همراه دلیل تغییر نسبت به پیش پردازش اصلی بیان خواهد کرد.

داده کاوی راه‌حلی برای به دست آوردن اطلاعات از میان این کوه‌های عظیم داده است؛ گسترش چشمگیر فناوری‌های کامپیوتری و در نتیجه افزایش حجم داده‌ها اهمیت این شاخه از هوش مصنوعی را بیش از پیش نموده است. روش‌های داده کاوی بیش‌تر بر روی داده‌های ساخت یافته مانند جداول و ماتریس‌ها متمرکز هستند؛ و زمانی که صحبت از کشف دانش از داده‌های متنی و ساخت نیافته‌ای مانند متن قرآن و یا ترجمه آن می‌شود حوزه‌ی متن کاوی (Text Mining) وارد عمل می‌شود. متن کاوی یا کشف دانش از متن به فرآیند استخراج دانش از متن اشاره دارد؛ روش‌های استخراج اطلاعات در متن کاوی از پردازش زبان طبیعی (Natural Language Processing) استفاده و آن را با الگوریتم‌ها و روش‌های داده کاوی، یادگیری ماشین (Machine Learning) و آمار مرتبط می‌کند. [۱]

آنچه که در یک فرآیند کامل متن کاوی صورت می‌پذیرد در سه گام انجام می‌شود؛ میزان دقت و صحت عمل در این سه گام مشخص کننده‌ی کیفیت نهایی فرآیند متن کاوی هستند. [۱۴]

۱. پیش پردازش (Pre Processing): در این گام آماده سازی داده‌ها برای تزریق به الگوریتم‌های یادگیری ماشین، مرتب سازی داده‌ها، برچسب گذاری داده‌ها صورت می‌پذیرد. هر یک از الگوریتم‌های یادگیری ماشین داده‌ها را در قالب ساختار یافته خاصی برای پردازش قبول می‌کنند و گاهی اوقات داده‌های خام ما متشکل شده از بخش‌های اضافی و یا بخش‌های اشتباهی هستند که برای به دست آمدن نتیجه مطلوب باید حذف گردند؛ نمونه‌ای این داده‌های اضافی داده‌های پرت (Outliers) هستند که در فاصله‌ای زیاد از اکثریت داده‌ها قرار دارند.

۲. کشف الگوهای داده (Modeling): در این گام با تزریق داده‌ها به الگوریتم‌های یادگیری ماشین و یا استفاده از شبکه‌های عصبی رابطه‌های میان داده‌ها را کشف می‌کنند. در این گام با کم و زیاد کردن متغیرهایی با عنوان Hyper parameter و یا بازگشت به گام قبل و ایجاد تغییر در داده‌ها کیفیت مدل‌های آموزش دیده را افزایش می‌دهند. داده‌هایی که در این گام به الگوریتم تزریق می‌شوند نمونه‌ها و مصادیق خروجی مطلوب ما هستند، یعنی هم داده و هم مقدار صحیحی که توقع داریم در آینده مدل موارد مشابه آن را برای ما پیشبینی کند را به الگوریتم می‌دهیم و الگوریتم پس از یادگیری تبدیل به مدل آموزش دیده می‌شود.

• برای ارزیابی میزان دقت مدل بخشی از داده‌ی اولیه که همراه با پاسخ درست است (داده برچسب گذاری شده) را که قبلاً به مدل تزریق نشده را به مدل تزریق می‌کنند و نتایجی که مدل پیشبینی می‌کند را با پاسخ‌های صحیح می‌سنجند، با نتایج به دست آمده ماتریس اغتشاش (Confusion Matrix) آزمایش را رسم می‌کنند.

۳.۵ استخراج داده

۳.۵.۱ استخراج داده با استفاده از هشتگ

برای استخراج توییت‌های حاوی هشتگ خاص توییت‌ر API منحصر به فردی را قرار داده تا به سادگی به تمامی توییت‌هایی که با این هشتگ ثبت شده اند دسترسی پیدا کنید. آنچه که این حالت از استخراج را حائز اهمیت قرار داده است، امکان ایجاد هشتگ اختصاصی برای پاسخ به شبهات و یا سوالات شرعی می‌باشد. در آزمایش مربوط به این پژوهش یک هشتگ اختصاصی منحصر به فرد با عنوان #آزمایش_پاسخ_شبهات در نظر گرفته شد که ربات استخراج گر بلافاصله پس از توییت شدن مطلبی حاوی این هشتگ آن را به پایگاه داده اضافه می‌کرد.

این روش بخصوص زمانی بسیار کارآمد است که هشتگ قراردادی مورد تبلیغ قرار گیرد و افرادی که واقعا شبهه‌ای برایشان به وجود آمده است و قصد گمراه کردن افراد دیگر را ندارد از آن استفاده کنند. در این روش پس از استخراج شدن توییت فقط لازم است دسته‌بندی مربوط به آن مشخص گردد تا به کارشناس مربوطه ارجاع گردد و پس از ثبت پاسخ توسط کارشناس محترم پاسخ بصورت کامنت روی توییت مدنظر ثبت گردد. در آزمایش این روش آنچه که از نظر عملیاتی موجب پیچیدگی برنامه نویسی شد، ارجاع مجدد بازخوردها به کارشناس مربوطه بود، یعنی یک سوال مطرح شده از سوی کاربران ممکن است با چندین بازخورد مواجه شود که این بازخوردها باید دقیقا توسط همان کارشناس قبلی پاسخ داده شوند چرا که احاطه فکری بر سوال و جواب درج شده دارد. [۹] برای رفع این مورد احتیاج به استفاده از پایگاه داده‌ای جدا از پایگاه داده اصلی بود، چرا که پایگاه داده حاوی توییت‌ها و پاسخ‌ها به دلیل انبوه اطلاعات MongoDB بود تا توانایی کار با داده‌های حجیم (Big Data) را داشته باشد. [۲] البته بدیهی است به علت منحصر به فرد بودن این هشتگ هنگام آزمایش تعداد توییت‌ها زیاد نبود، اما این روش زمانی بازدهی اصلی را دارد که به عنوان یک ابزار رسیدن به پاسخ شبهات به جامعه معرفی گردد.

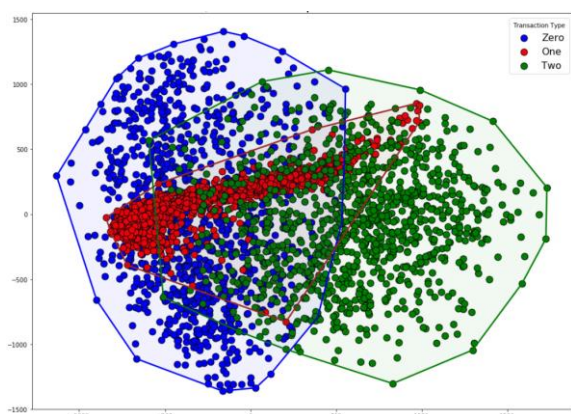
۳.۵.۲ استخراج داده با استفاده از کلیدواژه

روش دیگری که با استفاده از آن اطلاعات از توییت‌ر استخراج می‌شود، جستجوی توییت‌های حاوی کلیدواژه‌های خاص می‌باشد. در این روش حجم داده‌های به دست آمده بسیار زیاد بوده و نیاز به پردازش توییت‌ها برای شناسایی و دسته‌بندی آن‌ها به مراتب بیشتر از روش اول می‌باشد. در واقع در این روش ربات استخراج گر به دنبال توییت‌ها می‌گردد میان توییت‌هایی که حاوی کلمات پرتکرار در شبهات هستند.

برای آزمایش این روش، پس از جستجوی کلید واژه خدا از میان ۱۶۷۸۰ توییت استخراج شده، دسته‌بندی دستی برای فراهم سازی نیاز مدل‌های تفکیک توییت‌های حاوی شبهات انجام شد. از میان ۱۵۰۰ توییت که مورد برچسب گذاری قرار گرفت، تعداد ۱۷۰ توییت حاوی سوالات و یا شبهاتی در مورد کلید واژه جستجو شده یعنی خدا بود. به طور تقریبی این تعداد توییت پیرامون کلیدواژه خدا و با این شرایط در حدود ۵ و نیم روز میان فارسی زبانان تولید شده بود. چگونگی تشخیص توییت‌های حاوی شبهه از دیگر موارد در بخش مربوطه به طور کامل بیان خواهد شد، اما آنچه که بیش از همه در این مرحله زمانبر و نیازمند صرف انرژی بود برچسب گذاری داده‌ها بود که البته برای یک پروژه با گستردگی بیشتر توجه به این نسبت‌ها احتیاج به حداقل برچسب گذاری ۳۰۰۰۰ داده خواهیم داشت تا مدل نهایی تفکیک گر توییت‌ها از دقت قابل اطمینان تری برخوردار باشد.

۳.۶ تفکیک توییت‌های حاوی شبهه

جهت تفکیک توییت‌های حاوی شبهه سه الگوریتم یادگیری ماشین متناسب با این کاربرد انتخاب شد، اولین الگوریتم Random Forest بود که دقتی معادل ۸۷٪ برای پیش بینی درست شبهات را به ارمغان آورد، این دقت پس از تغییر ضرایب بطور اتوماتیک توسط Grde Search به دست آمد و به طور تقریبی فرآیند جستجوی بهینه‌ترین Hyper Parameters برای الگوریتم و آموزش بهترین مدل ۱۷ ساعت به طول انجامید. دومین و سومین الگوریتم‌ها به ترتیب K-Nearest Neighbors و Support vector machines بود که دقت‌هایی کمتر از ۸۰ درصد را به دست آوردند.



تصویر ۲: نمودار پراکنندگی تفکیک

همچنین باید اشاره نمود که پس از کاهش ابعاد با استفاده از الگوریتم PCA برای حصول اطمینان از دقت و نمایش خروجی روی نمودار، مطابق تصویر ۲ داده‌ها به سه دسته که مطابق آزمایش برای سه حالت شبهه بودن رنگ سبز و شبهه نبودن بطور قطعی رنگ آبی و همچنین توییت‌هایی که به طور قطعی نمی‌توان در مورد آن‌ها نظر داد به علت پایین بودن ضریب نزدیک بودن به هر دو دسته آورده شده است. این

کمترین در دیتاست مورد بررسی قرار گرفته، مهدویت، امامان، خدا، قرآن بود.

۳.۸ اتوماسیون سازی

آنچه که تا به اینجای کار بررسی شد، نحوه انجام هر مرحله به صورت دستی بود. اما قطعاً در یک پروژه با کاربری واقعی احتیاج به اتوماسیون سازی تمام مراحل داریم، به نحوی که تمامی این فرآیندها به صورت ترتیبی و بدون دخالت انسانی تا جایی که احتیاج به حضور اپراتور باشد به صورت خودکار فعالیت انجام دهد. چگونگی انجام این کار با استفاده از زبان برنامه نویسی پایتون و برخی از تکنولوژی‌های مدیریت صف بیان می‌شود. برای آزمایش این کار در اولین مرحله نرم افزار Apache Kafka بود که از نسخه ۳.۰ استفاده شد. این برنامه وظیفه مدیریت داده‌ها به صورت یک صف را بر عهده دارد، این قابلیت برای ایجاد چرخه‌های داده بسیار کاربردی می‌باشد. [۶] در اولین مرحله کتابخانه kafka-python نصب شد که برای دریافت و یا ارسال داده بر بستر این برنامه با استفاده از زبان برنامه نویسی پایتون می‌باشد. سپس اولین مرحله استخراج داده بود که پس از استخراج داده‌ها بر بستر اولین خط داده منتشر می‌شدند.

در طرف دیگر بخش مربوط به تشخیص شبهات بود که در صورتی که یک توییت حاوی شبهه بود به بخش بعدی یعنی دسته بندی از طریق همین ابزار منتقل می‌شد. سپس در آخرین مرحله و دسته‌بندی متناسب با دسته‌ها، شبهه به همراه اطلاعات مربوط به توییت ارسال شده به دیتابیس منتقل می‌شد تا کارشناس مربوطه پس از پاسخ به این توییت، مستقیماً به صورت کامنت بر روی توییت اصلی قرار گیرد. این حالت از اتوماسیون سازی مستلزم ایجاد رویه‌های تحمل خطا می‌باشد. البته باید اشاره کرد که برای انجام همین فرآیند می‌توان از نرم‌افزار Redis نیز استفاده کرد که عملکردی مشابه Kafka در ایجاد صف‌ها دارد، اما برخی امکانات بیشتر که در این زمینه کمتر مورد استفاده قرار می‌گیرند نیز به همراه خواهد داشت.

۴ بحث و نتیجه‌گیری

۴.۱ مقدمه

یافته‌های این پژوهش به دو دسته تقسیم می‌شوند، دسته اول مسیر ایجاد یک اتوماسیون پاسخ به شبهات توییت را نشان می‌دهند و دسته دوم نتایج آماری که در آزمایش‌های این پژوهش به دست آمد را به همراه دارند. هدف اصلی این پژوهش دسته اول بود، اما یافته‌های دیگر نیز در قالب مزیت‌هایی که این ایده علاوه بر اهداف اصلی ارائه می‌دهند مورد بررسی قرار خواهد گرفت.

مقدار زیاد از داده‌هایی با دسته بندی نامعلوم به علت فیلتر گذاری برای ضرایب بالای ۸۰٪ به وجود آمده است و به مرور زمان با به دست آمدن دیتاست‌های دقیق تر تعداد داده‌های مجهول نیز کاهش می‌یابد. این مرحله از پروژه مربوط به این پژوهش پیش‌پردازش، آموزش مدل یادگیری ماشین و آزمایش مدل را در بر داشت، در این فاز برای مشاهده تاثیر تغییرات بر روی مدل مطابق اصول CRISP چندین بار جهش بین این مراحل انجام گردید تا در نهایت مدل به حالتی رسید که با استفاده از Grice Search قابلیت بهینه سازی پیدا کرد.

باید اشاره داشت که در این مثال با استفاده از الگوریتم Random Over-Sampling دیتاست استاندارد سازی و Balance شد تا خطای ناشی از کم بودن پراکندگی یکی از دسته‌های تفکیکی کاهش یابد. نسبت تعداد دسته‌های مختلف یکی از عوامل مهم در مسئله‌های دسته بندی است، چرا که کم بودن تعداد یکی از دسته‌ها منجر به Overfit مدل سمت داده‌های بیشتر و Underfit شدن داده سمت داده‌های کمتر می‌شود. برای کنترل این موضوع هم افزودن داده به کمک الگوریتم مورد استفاده در این مسئله و هم کاهش تعداد دسته با تعداد بیشتر با الگوریتم Random Under-Sampling امکان پذیر است. باید دقت داشت که هر الگوریتم بسته به روش‌های متفاوتی که به کار می‌گیرد برای افزایش یا کاهش داده، مثلاً محل انتخابی برای افزایش و یا کاهش داده، از کاربرد متفاوتی نیز برخوردار است. [۲]

۳.۷ دسته بندی شبهات

برای اینکه هر شبهه بطور تخصصی توسط کارشناس خاصی پاسخ داده شود و گزارش دقیقی از محور شبهات به دست آید، باید توییت‌ها در آخرین مرحله قبل از رسیدن به دست کارشناس مربوطه مورد دسته‌بندی قرار گیرند. در این مرحله برای دسته بندی از الگوریتم Support vector machines استفاده شد. پس از آزمایش الگوریتم Random Forest و همچنین گزارش‌های قبلی در مورد این الگوریتم [۵] زمانی که تعداد دسته‌ها برای دسته‌بندی افزایش یابد، سرعت کاهش یافته و در مقایسه با برخی از الگوریتم‌ها زمان بیشتری احتیاج دارد.

پس از مشاهده نتیجه آزمایش، مشخص شد که پیش‌پردازش مربوط به بخش قبلی برای این الگوریتم نیز از سازگاری بالایی برخوردار است. برای دسته‌بندی ۴ محور اصلی شبهات انتخاب گردید. خدا، مهدویت، قرآن و امامان که در این دسته‌ها مورد دسته‌بندی قرار گیرند. البته این دسته‌بندی‌ها برای پروژه‌های بزرگ‌تر قابل افزایش و یا کاهش هستند. سپس برای فرآیند برچسب‌گذاری تعداد ۱۵۰۰ عدد از توییت‌ها مورد برچسب گذاری قرار گرفتند و از دیتاست حاصله برای دسته‌بندی استفاده شد. پراکندگی این شبهات به ترتیب بیشترین به

اصلی ترین سوال های این پژوهش که محور آزمایشات و تحقیقات بر مبنای آن ها نباشد، امکان پذیر بودن یا نبودن استخراج شبهات منتشر شده در توییتر به صورت اتوماتیک و چگونگی این عمل بود. یافته های این پژوهش نشان می دهد که استخراج این شبهات با استفاده از داده کاوی با استفاده از مراحل بیان شده در روش پژوهش به خوبی پاسخگوی نیاز ما به استخراج شبهات به صورت هدفمند می باشد. در این پژوهش برای پاسخ به این سوال از داده کاوی و هوش مصنوعی استفاده شد. یعنی از حوزه ها و الگوریتم هایی که این علم در اختیار ما قرار داده است استفاده شد تا به پاسخ مطلوب برسیم. بیشترین حوزه هایی که در این پژوهش مورد استفاده قرار گرفتند، به ترتیب یادگیری ماشین و متن کاوی می باشد که در محل مربوطه چستی و چگونگی بهره بردن از آن ها در راستای این پژوهش بطور خلاصه بیان شد.

نکته کلیدی در بکارگیری این پژوهش برای پیاده سازی یک پروژه جدید در این راستا، طراحی Pipeline متناسب به حجم کاری مورد نیاز می باشد. یعنی برای پروژه های که مانند یک سرویس دائمی احتیاج به استفاده از این روش دارد باید به این نکته توجه داشت که باید فرآیند سازی صورت پذیرد و فرآیندها با استفاده از تکنولوژی های برنامه نویسی به یکدیگر بدون وابستگی زمانی به مانند یک صف پشت سر یکدیگر قرار گیرند. در آخرین مرحله از این چرخه کارشناسان پاسخگو قرار می گیرند که توسط یک رابط کاربری فقط به متن توییت های حاوی شبهه دسترسی دارند و به آن پاسخ می دهند. [۳]

۴.۳ یافته های کاربردی

از جمله کاربردی ترین یافته های این پژوهش مطالب مربوط به بخش مقدمه روش پژوهش می باشد که اشاره به محدودیت های توییتر در دسترسی به داده ها می باشد و پس از آن چگونگی اتوماسیون سازی این فرایند با استفاده از ابزارهای معرفی شده که برای پیاده سازی یک سیستم هوشمند بر پایه توییتر کاربردی و کلیدی است. دیگر یافته کاربردی این پژوهش آمارهای شبهات می باشد که تفصیل دوباره بیان می شوند. اولین مورد در مورد نتایج آماری این پژوهش نسبت های تعداد شبهات می باشد که بیشترین متعلق به شبهات مربوط به مهدویت می باشد. این امر نشانگر پرسش ها و کنجکاوی بیشتر جامعه حاضر در توییتر که نمونه ای از جامعه واقعی فارسی زبان می باشد نسبت به مفاهیم مهدویت و اشتیاق به کسب اطلاعات بیشتر در این زمینه می باشد. البته باید دقت داشت که در بازه های زمانی مختلف به سبب شرایط جامعه این نتایج مورد تاثیر قرار گرفته و دستخوش تغییر می شوند. به سبب این موضوع باید تحلیل ها به صورت دوره ای در این زمینه تکرار و یا اتوماسیون سازی شوند،

همچنین از این روش برای سایر موارد مورد نظر در هر حوزه ای می توان استفاده نمود و از داده های استخراج شده به جهت پیشگیری و یا راه حل استفاده کرد.

۵ پیشنهادها

۵.۱ مقدمه

آنچه که در این پژوهش مورد بررسی قرار گرفت، حاصل نگاه صنعتی تر و آمیزه ای از تجاری نگری به حوزه علوم دینی بود. یعنی از ابزارهایی که کسب و کارها برای گسترش خود استفاده می کنند در حوزه گسترش علوم دینی بهره های فراوانی می توان برد که احتیاج به اهمیت دادن و تمرکز بیشتر بر این موضوع دارد. [۱۲]

۵.۲ بررسی پیشنهادها

۱. بسیار کاربردی و مفید می باشد که برخی از پژوهش ها در قالب یافته های آماری مصور سازی شده به جامعه عرضه شود با این عنوان که در یک دوره خاص تاریخی در زمینه دینی و مذهبی مردم یک جامعه چه گرایشاتی داشته اند و این موضوع بر اساس تحلیل بزرگان به چه دلایلی می تواند باشد.

۲. استفاده از ابزارهای هوش مصنوعی و آمار برای تحلیل واکنش مردم نسبت به موضوع دینی خاص و مانور رسانه ای و یا بهبود آن موضوع. برای مثال از طریق شبکه های اجتماعی میزان نارضایتی مردم از مسائل مختلفی نظیر ربا، بی عدالتی اجتماعی در حوزه قانون اساسی اسلامی و محدودیت های فردی و اجتماعی جوامع اسلامی صورت پذیرد و بر اساس میزان نارضایتی مردم به ترتیب اولویت برنامه های توجیهی و یا اصلاحی برای موارد فوق بر بستر شبکه های اجتماعی بخصوص صورت پذیرد.

۳. پژوهش بیشتر در زمینه تفسیر برای ایجاد تفسیر هایی مطابق با زبان و فهم مخاطب، برای مثال تفسیر هایی مناسب سن و جنس خاص (از نظر بیانی و نزدیکی به درک مخاطب) برای توسعه بیشتر فرهنگ قرآنی و تمدن اسلامی میان مسلمانان.

۴. تحقیق و توسعه بیشتر در زمینه NLP و اتوماتیک سازی فرآیند پیش پردازش با استفاده از نیازسنجی الگوریتم های متن کاوی مشهور.

۵. ایجاد زمینه های یادگیری معکوس در زمینه متن کاوی به همراه تاریخ شناسی زمان نزول قرآن کریم (فرآیند شناسی مفاهیم و آموزه ها) با استفاده از توسعه و گسترش دادن پژوهشگاه های قرآنی میان رشته ای.

۶. در حال حاضر پروژه ها و تحقیقاتی وجود دارند که متن را تبدیل به تصویر می کنند، در زمینه قرآن کریم با توجه به وجود

داستان های بسیار و توصیف های بی بدیل از مکان های خاص نظیر بهشت و جهنم این فرصت وجود دارد که با توسعه هوش مصنوعی در این زمینه اقدام به تصویرسازی های نزدیکی برای انزار بیشتر به پیروان صورت پذیرد.

۷. بهره بردن از هوش مصنوعی برای صحت سنجی روایات و احادیث دینی و مذهبی، بر اساس واکاوی و مفهوم شناسی منابع موجود و اسناد آورده شده. این مهم از این حیث حائز اهمیت است که پیش زمینه رصد و تحلیل تفکرات اشتباه و احادیث ضعیف از نظر سند و منبع را فراهم می کند و سپس با استفاده از خروجی باید یک برنامه دقیق برای محو سازی و جایگزین سازی موارد اشتباه با موارد صحیح صورت پذیرد.

۶ مراجع

- [۱] Al-Amin, S. T., & Ordonez, C. (2021). Efficient machine learning on data science languages with parallel data summarization. *Data & Knowledge Engineering*, 136, 101930. <https://doi.org/10.1016/j.datak.2021.101930>
- [۲] Data Processing through Data Warehouse and Data mining. (2017). *International Journal of Modern Trends in Engineering & Research*, 4(5), 45–48. <https://doi.org/10.21884/ijmter.2017.4151.1ea3x>
- [۳] Kalokhe, Prof. P. (2017). Obtaining Good Heuristics for Big Data Broadcasting Problem Using Novel Pipeline Approach. *International Journal Of Engineering And Computer Science*. <https://doi.org/10.18535/ijecs/v6i5.01>
- [۴] Lee, E.-J., Lee, H.-Y., & Choi, S. (2020). Is the message the medium? How politicians' Twitter blunders affect perceived authenticity of Twitter communication. *Computers in Human Behavior*, 104, 106188. <https://doi.org/10.1016/j.chb.2019.106188>
- [۵] Mantas, C. J., Castellano, J. G., Moral-García, S., & Abellán, J. (2018). A comparison of random forest based algorithms: Random credal random forest versus oblique random forest. *Soft Computing*, 23(21), 10739–10754. <https://doi.org/10.1007/s00500-018-3628-5>
- [۶] Shaheen, J. A. (2017). Apache Kafka: Real Time Implementation with Kafka Architecture Review. *International Journal of Advanced Science and Technology*, 109, 35–42. <https://doi.org/10.14257/ijast.2017.109.04>
- [۷] اعرافی، علیرضا. (۱۴۰۰). مهم ترین راهبردهای پاسخگویی به شبهات. پاسخ، ۲۳(۶)، ۷–۱۴.

تحقیقات اجتماعی در ایران.

- [۹] زمانی، محمدحسن. (۱۴۰۰). بایسته های مهارت پاسخگویی به شبهات دینی. پاسخ، ۲۲(۶)، ۷–۱۸.
- [۱۰] صابر، زینب دانشگاه شیراز، یوسفی، زهرا دانشگاه خلیج فارس - گروه علم اطلاعات و دانش شناسی، خونجوش، فرشاد دانشگاه شیراز - بخش مهندسی و علوم کامپیوتر و فناوری اطلاعات، & قنبری آلونی، فرزین دانشگاه شیراز. (۱۳۹۹). تحلیل محتوا و تحلیل عقاید توییت ها درباره دسترسی آزاد و جنبه های مختلف آن. پژوهش نامه پردازش و مدیریت اطلاعات.
- [۱۱] روحانی، سعید دانشگاه تهران - گروه مدیریت فناوری اطلاعات. (۱۳۹۸). طراحی و آزمایش داشبورد بلادرنگ تحلیل متن شبکه اجتماعی توییت. پردازش علائم و داده ها.
- [۱۲] عابدینی، حسین نویسنده دانشگاه صنعتی شریف. (n.d.). کاربردهای داده کاوی در علوم اسلامی. ره آورد نور.
- [۱۳] هاشمی، ضیاء دانشگاه تهران - دانشکده علوم اجتماعی، & رهبری، محمد دانشگاه علامه طباطبائی تهران. (۱۳۹۹). شبکه های اجتماعی مجازی و قبح زدایی از محتوای مبتذل (مطالعه موردی: توییت فارسی). مطالعات رسانه های نوین.
- [۱۴] رضائیان مینا و همکاران. (۱۳۹۵). «متن کاوی ابزاری مفید در آینده پژوهی: معرفی روش و مرور ادبیات» سومین همایش ملی آینده پژوهشی؛

- [۸] جلائی پور، حمیدرضا دانشگاه تهران - گروه علوم اجتماعی & حاجی زادگان، ابوالفضل دانشگاه تهران. (۱۳۹۹). شناسایی و ارزیابی الگوهای تأثیرپذیری کاربران ایرانی توییت از نگرش های سیاسی اینفلوئنسر ها. مطالعات و