تخمین کیفیت پاسخ با استفاده از یادگیری جمعی در جوامع پرسش و پاسخ

معصومه خالقیان[۱*] و میرمحسن پدرام[۲]

[۱]دانشجوی مقطع کارشناسی ارشد رشته هوش‌مصنوعی از دانشگاه خوارزمی، masoumehkhaleghian@khu.ac.ir

[۲]دانشیار گروه مهندسی برق و کامپیوتر دانشکده فنی برق خوارزمی، pedram@khu.ac.ir

چکیده : جوامع پرسش و پاسخ به بخشی جدایی‌ناپذیر از زندگی روزمره انسان‌ها تبدیل شده‌اند. با وجود موفقیت این پلتفرم‌ها، بسیاری از پرسش‌ها بدون انتخاب پاسخ بهینه باقی می‌مانند، زیرا پرسش‌گران هیچ پاسخی را به‌عنوان بهترین گزینه تأیید نمی‌کنند. این وضعیت منجر به اتلاف زمان کاربران در جستجوی پاسخ‌های مناسب می‌شود. با رشد مداوم این جوامع، شناسایی دقیق بهترین پاسخ‌ها به چالشی اساسی تبدیل شده است. در این پژوهش، یک مدل یادگیری جمعی وزن‌دار ارائه شده است که چهار یادگیرنده پایه قدرتمند شامل *XLNet*، *RoBERTa*، جنگل‌های تصادفی و رأی‌گیری اکثریتی را ترکیب می‌کند. فرایادگیر این مدل که مبتنی بر درخت تصمیم است، با وزن‌دهی مناسب به خروجی یادگیرنده‌های پایه، دقت مدل را به سطح بی‌نظیر ۹۹.۰۳٪ در مجموعه‌داده آزمون ارتقا داده است. نتایج نشان می‌دهند مدل پیشنهادی می‌تواند نقش مهمی در بهبود کیفیت پاسخ‌ها، افزایش اعتماد به اطلاعات، و ارتقای تجربه کاربری در سیستم‌های پرسش و پاسخ ایفا کند.

کلید واژه‌ها: تشخیص بهترین پاسخ، یادگیری جمعی، مدل‌های زبانی، پردازش زبان طبیعی، سیستم‌های پرسش و پاسخ

۱- مقدمه

با رشد سریع جوامع پرسش و پاسخ و افزایش چشمگیر حجم اطلاعات، نیاز به سیستم‌های هوشمند برای تشخیص خودکار بهترین پاسخ‌ها بیش از پیش احساس می‌شود. در عصر دیجیتال، تعاملات انسانی به طور گسترده به فضاهای آنلاین منتقل شده است و یکی از مهم‌ترین این فضاها، جوامع پرسش و پاسخ آنلاین هستند. این پلتفرم‌ها، با ایجاد بستری برای طرح پرسش‌ها و دریافت پاسخ‌های تخصصی، به ابزاری قدرتمند برای حل مسائل و ارتقای دانش فردی و جمعی تبدیل شده‌اند.

با این حال، این جوامع با چالش‌هایی نظیر حجم عظیم اطلاعات، تنوع موضوعات و کیفیت متغیر پاسخ‌ها مواجه هستند که می‌توانند کارایی آن‌ها را محدود کنند. در چنین شرایطی، بهره‌گیری از روش‌های پیشرفته هوش مصنوعی و یادگیری ماشین به عنوان راهکاری مؤثر مطرح می‌شود.

این پژوهش با هدف بهبود فرآیند شناسایی بهترین پاسخ‌ها و ارتقای کارایی این پلتفرم‌ها انجام شده است. در این راستا، مدل‌های قدرتمند زبانی نظیر **XLNet** و **RoBERTa** برای پردازش ویژگی‌های متنی و مدل‌هایی همچون جنگل‌های تصادفی برای تحلیل ویژگی‌های غیرمتنی مورد استفاده قرار گرفته‌اند. همچنین، ترکیب این مدل‌ها از طریق روش‌های پشته‌ای و استفاده از فرایادگیر مبتنی بر درخت تصمیم، امکان دستیابی به دقتی فراتر از توانایی مدل‌های منفرد را فراهم کرده است.

این تحقیق تلاش دارد با ارائه مدل‌های ترکیبی پیشرفته و بهبود دقت پیش‌بینی‌ها، چالش‌های موجود در جوامع پرسش و پاسخ آنلاین را کاهش داده و به ارتقای کیفیت اطلاعات، افزایش رضایت کاربران و بهبود تجربه کاربری کمک کند.

۲- بیان مسأله و اهداف پژوهش

پلتفرم‌های پرسش و پاسخ آنلاین بستری را فراهم می‌کنند که کاربران بتوانند دانش خود را در حوزه‌های مختلف به اشتراک بگذارند و به دنبال پاسخ‌های مرتبط با سوالات خود باشند. این پلتفرم‌ها به دلیل توانایی‌شان در ارائه پاسخ‌های تخصصی و متناسب با نیاز کاربران، به یکی از منابع اصلی اطلاعات تبدیل شده‌اند. با این حال،

چالش‌های متعددی در این زمینه وجود دارد، از جمله حجم زیاد اطلاعات، کیفیت متغیر پاسخ‌ها و دشواری در شناسایی پاسخ‌های بهینه.

مسئله اصلی این است که کاربران غالباً نمی‌توانند به‌راحتی بهترین پاسخ ممکن را از میان تعداد زیادی از پاسخ‌های ارائه‌شده پیدا کنند. این موضوع منجر به صرف زمان زیاد و کاهش کارایی در استفاده از این جوامع می‌شود. علاوه بر این، معیارهای ارزیابی پاسخ‌ها به دلیل تأثیرگذاری عوامل مختلف متنی و غیرمتنی همچنان موضوعی چالش‌برانگیز است.

در این پژوهش، هدف ارائه راهکاری مبتنی بر هوش مصنوعی است که بتواند با تحلیل ویژگی‌های متنی (مانند کیفیت محتوا) و غیرمتنی (مانند تعداد رأی‌های مثبت) بهترین پاسخ را به صورت خودکار شناسایی کند. تمرکز این مطالعه بر شناسایی عواملی است که یک پاسخ را به‌عنوان بهترین پاسخ برای کاربران تبدیل می‌کند، به‌طوری‌که نیاز اطلاعاتی آن‌ها را به بهترین شکل ممکن برآورده سازد.

۲-۱- ضرورت انجام پژوهش

در جوامع پرسش و پاسخ، تعاملات کاربران و تنوع محتوای تولید شده پیچیدگی زیادی دارد و الگوریتم‌های معمول اغلب در مدل‌سازی این پیچیدگی‌ها و پاسخ‌گویی به نیازهای اطلاعاتی کاربران ناکارآمد هستند.

با افزایش استفاده از پلتفرم‌های پرسش و پاسخ آنلاین، این فضاها به یکی از منابع اصلی برای جستجوی اطلاعات در موضوعات متنوع تبدیل شده‌اند. با این حال، حجم بالای داده‌ها و تنوع کیفیت پاسخ‌ها، چالشی جدی برای کاربران به شمار می‌رود. شناسایی و انتخاب بهترین پاسخ در میان حجم انبوهی از اطلاعات موجود، نه تنها زمان‌بر است بلکه گاهی منجر به تصمیم‌گیری‌های نادرست می‌شود.

اهمیت پژوهش حاضر در این است که بتوان با بهره‌گیری از روش‌های هوش مصنوعی و تحلیل ویژگی‌های متنی و غیرمتنی پاسخ‌ها، به کاربران در یافتن پاسخ‌های بهینه کمک کرد. این کار نه تنها باعث افزایش کارایی و دقت در شناسایی پاسخ‌های مناسب می‌شود، بلکه تجربه کاربری را نیز بهبود می‌بخشد. از سوی دیگر، با توجه به نقش مهم اطلاعات در تصمیم‌گیری‌های روزمره و حرفه‌ای، ارائه روشی کارآمد برای بهبود دسترسی به پاسخ‌های باکیفیت، ضرورتی غیرقابل انکار است.

علاوه بر این، توسعه مدلی که بتواند ویژگی‌های متنی و غیرمتنی را به صورت هم‌زمان تحلیل کند، می‌تواند به پیشرفت‌های قابل توجهی در حوزه‌های مرتبط با پردازش زبان طبیعی و یادگیری ماشین منجر شود. این پژوهش با تمرکز بر داده‌های واقعی و روش‌های نوین، می‌تواند به‌عنوان پایه‌ای برای توسعه سیستم‌های توصیه‌گر در جوامع پرسش و پاسخ عمل کرده و نقش مهمی در بهبود تعاملات کاربران ایفا کند.

۲-۲- چالش‌های موجود

با وجود رشد روزافزون پلتفرم‌های پرسش و پاسخ آنلاین و اهمیت عملیاتی این سیستم‌ها، پیش‌بینی خودکار بهترین پاسخ همچنان با چالش‌های متعددی روبرو است. این چالش‌ها از ویژگی‌های ساختاری و مفهومی این جوامع سرچشمه می‌گیرند و نیازمند رویکردهایی پیشرفته و ترکیبی برای حل مسائل موجود هستند.

یکی از چالش‌های اساسی، **نبود پاسخ‌های واضح و قطعی** برای بسیاری از سوالات است. در بسیاری از موارد، پاسخ‌دهندگان از دیدگاه‌ها و رویکردهای مختلف به سوال پاسخ می‌دهند، که این موضوع انتخاب بهترین پاسخ را دشوار می‌کند. علاوه بر این، **تعداد زیاد شرکت‌کنندگان در بحث و تعاملات پیچیده میان کاربران** باعث می‌شود که ارزیابی و تحلیل کیفی پاسخ‌ها پیچیدگی بیشتری پیدا کند.

بسیاری از پژوهش‌های قبلی این مسائل را به طور کامل مدنظر قرار نداده‌اند. برای مثال، در سیستم‌های پرسش و

پاسخ حوزه باز، پاسخ‌ها اغلب در مجموعه‌داده‌های عمومی یا منابعی مانند ویکی‌پدیا قرار دارند که شامل پاسخ‌های غیرمتمرکز و نامحدود است. اما در سیستم‌های پرسش و پاسخ اجتماعی، ساختار ساده‌تر است؛ سوال توسط یک کاربر مطرح شده و اعضای جامعه به آن پاسخ می‌دهند. با این حال، نبود معیارهای مشخص و جامع برای ارزیابی کیفیت پاسخ‌ها همچنان یک معضل کلیدی است.

یکی دیگر از چالش‌های برجسته، **فرضیات ساده‌سازی شده در مدل‌های کنونی** است. این مدل‌ها غالباً فرض می‌کنند که کاربران سوال‌کننده دانش و دیدگاه مشابهی با پاسخ‌دهندگان دارند، در حالی که در عمل، تفاوت‌های قابل توجهی در سطح دانش، باورها و نیازهای اطلاعاتی کاربران وجود دارد. مدل‌های فعلی به دلیل نادیده گرفتن این تنوع، قادر به در نظر گرفتن تفاوت‌های مهم در فرایند انتخاب بهترین پاسخ نیستند.

چالش‌های کلیدی در پیش‌بینی بهترین پاسخ به شرح زیر هستند:

- **کمبود داده‌های برچسب‌خورده باکیفیت:** بسیاری از مجموعه‌داده‌های موجود فاقد اطلاعات کافی برای آموزش مدل‌های پیشرفته یادگیری ماشین هستند.
- **کیفیت پایین داده‌ها:** وجود داده‌های نویزی و پاسخ‌های بی‌کیفیت می‌تواند دقت مدل‌ها را به شدت کاهش دهد.
- **تعاملات پیچیده میان کاربران:** روابط متقابل کاربران، مانند رای‌دهی و بازخورد، نیازمند مدل‌سازی دقیق‌تری هستند.
- **تغییرات زمانی:** کیفیت و اهمیت پاسخ‌ها ممکن است در گذر زمان تغییر کند و مدل‌ها باید این تغییرات را منعکس کنند.
- **تنوع زبانی و فرهنگی:** تفاوت در شیوه بیان و درک مفاهیم میان کاربران با پیشینه‌های مختلف می‌تواند فرایند پیش‌بینی را پیچیده‌تر کند.
- **سوگیری کاربران:** نظرات ذهنی کاربران و الگوهای رفتاری آن‌ها می‌تواند به نتایجی نامعتبر منجر شود.
- **پاسخ‌های چندمعنایی و متناقض:** وجود پاسخ‌هایی با تفسیرهای مختلف و گاه متناقض، شناسایی پاسخ بهینه را دشوار می‌سازد.
- **عدم تفسیرپذیری نتایج:** مدل‌های پیشرفته یادگیری ماشین گاهی فاقد شفافیت کافی برای توضیح انتخاب‌های خود هستند.

این چالش‌ها نشان می‌دهند که توسعه سیستم‌های پیشرفته‌تر برای پیش‌بینی بهترین پاسخ، نه تنها نیازمند استفاده از تکنیک‌های یادگیری ماشین و پردازش زبان طبیعی است، بلکه به ترکیب روش‌های چندگانه و بهره‌گیری از داده‌های متنی و غیرمتنی نیز نیاز دارد.

برای رفع این محدودیت‌ها، پژوهش حاضر تلاش دارد با ارائه رویکردی ترکیبی که ویژگی‌های زبانی، غیرمتنی و رابطه‌ای را در نظر می‌گیرد، دقت و کارایی انتخاب خودکار بهترین پاسخ را افزایش دهد. هدف نهایی این است که با توسعه مدل‌های هوشمند و قابل تفسیر، پلتفرم‌های پرسش و پاسخ آینده قادر باشند حتی در شرایط پیچیده و با وجود عدم قطعیت، پاسخ‌های بهینه را به کاربران ارائه دهند.

۳-۲- مفروضات و سوال‌های پژوهش

این پژوهش بر اساس مفروضاتی طراحی شده است که نقش اساسی در شکل‌گیری روش پیشنهادی ایفا می‌کنند. نخست آنکه، فرض بر این است که استفاده از رویکرد آموزش نظارت‌شده می‌تواند با کاهش پیچیدگی مدل، دقت و کارایی آن را در شناسایی بهترین پاسخ‌ها افزایش دهد. همچنین، انتخاب تابع زیان آنتروپی متقاطع باینری به‌عنوان معیار بهینه‌سازی، نقش مهمی در بهبود فرایند یادگیری و دقت در تشخیص پاسخ‌های بهینه ایفا می‌کند. داده‌های مورد استفاده در این پژوهش از مجموعه پرسش و پاسخ‌های مرتبط با حوزه اندروید در

پلتفرم StackOverflow، یکی از زیرمجموعه‌های شبکه StackExchange، جمع‌آوری شده است. این داده‌ها شامل ویژگی‌های متنی و غیرمتنی هستند که به صورت جداگانه استخراج و سپس ترکیب می‌شوند تا تأثیر استفاده از ویژگی‌های چندگانه در بهبود عملکرد مدل بررسی شود.

در راستای تحقق اهداف پژوهش، سؤالاتی مطرح می‌شود که جهت‌گیری تحقیق را مشخص می‌کنند. یکی از این پرسش‌ها این است که آیا رویکردهای کلاس‌بندی مبتنی بر برچسب‌دهی می‌توانند استراتژی مناسبی برای آموزش مدل در شناسایی بهترین پاسخ‌ها باشند. همچنین این سؤال بررسی می‌شود که آیا استفاده از یادگیری عمیق به صورت انتها به انتها می‌تواند استخراج ویژگی‌های متنی را به شکلی دقیق‌تر و اثربخش‌تر انجام دهد. از سوی دیگر، تأثیر ترکیب ویژگی‌های متنی و غیرمتنی بر دقت و کارایی مدل مورد توجه قرار می‌گیرد. در نهایت، این پژوهش بررسی می‌کند که آیا استفاده از مجموعه‌داده StackOverflow، به‌عنوان منبعی تخصصی در حوزه اندروید، می‌تواند در شناسایی دقیق‌تر پرسش‌ها و پاسخ‌های مرتبط مؤثر باشد.

٤-٢- نوآوری پژوهش

پیش‌بینی بهترین پاسخ در جوامع پرسش و پاسخ به عنوان یکی از مسائل مهم در حوزه هوش مصنوعی و یادگیری ماشین شناخته می‌شود. در این پژوهش، ما با استفاده از یک مدل یادگیری جمعی پشته‌ای وزن‌دار، تحولی جدید در دقت و کارایی این سیستم‌ها ایجاد کرده‌ایم. مدل پیشنهادی ما با ترکیب چندین روش مدرن و نوآورانه، توانسته است دقت پیش‌بینی را به میزان ۹۹.۰۳٪ افزایش دهد، که در مقایسه با پژوهش‌های پیشین، یک جهش چشمگیر محسوب می‌شود.

پژوهش حاضر با هدف ارائه راهکاری جامع و دقیق برای پیش‌بینی بهترین پاسخ‌ها در جوامع پرسش و پاسخ، مجموعه‌ای از رویکردهای نوآورانه و تکنیک‌های پیشرفته را به کار گرفته است.

٤-٢-١- **استفاده از مدل‌های پیشرفته زبانی**
این پژوهش از مدل‌های زبانی قدرتمند **XLNet** و **RoBERTa** برای استخراج خودکار ویژگی‌های متنی استفاده کرده است. این مدل‌ها با توانایی بی‌نظیر در درک معنایی عمیق و تحلیل پیچیدگی‌های متنی، نقش کلیدی در بهبود عملکرد مدل نهایی ایفا کرده‌اند.

٤-٢-٢- **بهره‌گیری از یادگیری جمعی وزن‌دار پشته‌ای**
مدل پیشنهادی با استفاده از تکنیک **استکینگ وزن‌دار** به شکلی کاملاً هوشمندانه، خروجی‌های مدل‌های پایه مانند XLNet، RoBERTa و جنگل‌های تصادفی را ترکیب کرده است. این روش از نقاط قوت هر مدل بهره‌برداری کرده و به شکل مؤثری ضعف‌های آن‌ها را پوشش داده است.

٤-٢-٣- **استفاده از خروجی درصد وقوع برای پیش‌بینی نهایی**
به جای استفاده از پیش‌بینی‌های قطعی صفر و یک، از خروجی درصد وقوع کلاس‌ها برای تصمیم‌گیری نهایی استفاده شده است. این رویکرد امکان تحلیل دقیق‌تر و انعطاف‌پذیری بیشتری در تعیین بهترین پاسخ‌ها فراهم کرده است.

٤-٢-٤- **ترکیب هوشمندانه ویژگی‌های متنی و غیرمتنی**
به جای استفاده از لایه‌های کاملاً متصل سنتی، ویژگی‌های متنی استخراج‌شده از مدل‌های زبانی و ۲۱ ویژگی غیرمتنی دست‌ساز به شکل هوشمندانه‌ای ترکیب شده‌اند. این ترکیب کارآمد باعث بهبود دقت مدل و بهره‌برداری بهتر از اطلاعات ساختاریافته و معنایی مجموعه‌داده شده است.

٤-٢-٥- **مقیاس وسیع داده‌ها**
این پژوهش از ۳۵,۵۶۸ جفت پرسش و پاسخ استخراج‌شده از

android.stackexchange.com استفاده کرده است. این حجم بالای داده‌ها، امکان آموزش و ارزیابی دقیق مدل را فراهم کرده و تعمیم‌پذیری نتایج را افزایش داده است.

۶-٤-۲- **دستیابی به دقت بی‌نظیر** مدل نهایی با دقت ۹۹.۰۳٪، توانسته است عملکردی فوق‌العاده در پیش‌بینی بهترین پاسخ‌ها ارائه دهد. این دقت نشان‌دهنده برتری مدل پیشنهادی نسبت به رویکردهای پیشین و توانایی بالای آن در حل مسئله است.

این نوآوری‌ها در کنار یکدیگر، پژوهش حاضر را به یک راهکار جامع و پیشرفته برای بهبود کیفیت تعاملات در جوامع پرسش و پاسخ تبدیل کرده‌اند.

۵-۲- مزایای کلیدی مدل پیشنهادی

مدل پیشنهادی با استفاده از ترکیبی از مدل‌های پایه قوی مانند XLNet، RoBERTa، جنگل‌های تصادفی و رأی‌گیری اکثریت طراحی شده است که هر کدام به‌طور خاص در تحلیل ویژگی‌های متنی و غیرمتنی تخصص دارند. این تنوع مدل‌ها نه تنها به یادگیری بهتر کمک کرده بلکه باعث افزایش انعطاف‌پذیری و بهبود عملکرد مدل نیز شده است. یکی از نوآوری‌های کلیدی این پژوهش استفاده از تکنیک وزن‌دار پشته‌ای[۱] است، جایی که وزن هر مدل پایه بر اساس دقت و عملکرد آن به‌دقت تنظیم شده است. این تکنیک با کاهش خطاهای مدل‌های پایه توانسته دقت پیش‌بینی نهایی را به طور چشمگیری افزایش دهد.

برای جلوگیری از بیش‌برازش[۲] و افزایش قابلیت تعمیم‌دهی مدل، از تکنیک اعتبارسنجی متقاطع[۳] استفاده شده است. این روش به‌ویژه در شرایطی که داده‌های برچسب‌خورده محدود هستند، نقش مؤثری در مدیریت این محدودیت ایفا کرده و عملکرد مدل را بهبود بخشیده است. همچنین، مدل پیشنهادی با بهره‌گیری از ترکیب ویژگی‌های متنی (مانند معنای عمیق متن) و غیرمتنی (مانند تاریخچه تعاملات کاربران) توانسته ارزیابی جامعی از کیفیت پاسخ‌ها ارائه دهد.

یکی دیگر از ویژگی‌های منحصربه‌فرد این مدل، استفاده از فرایادگیرنده درخت تصمیم است که با دریافت خروجی مدل‌های پایه و بهره‌گیری از درصد وقوع کلاس‌ها به جای پیش‌بینی کلاس دقیق، ترکیب هوشمندانه‌ای از پیش‌بینی‌ها ایجاد کرده است. این روش علاوه بر دقت بالا، انعطاف‌پذیری مدل را نیز افزایش داده است.

در نهایت، مدل پیشنهادی توانایی قابل‌توجهی در مدیریت داده‌های کم دارد. استفاده از اعتبارسنجی متقاطع و ترکیب هوشمندانه مدل‌های پایه به کمک فرایادگیرنده، این امکان را فراهم کرده است که مدل در شرایط واقعی و با داده‌های محدود نیز عملکرد مناسبی داشته باشد. این ویژگی‌ها مدل پیشنهادی را به یک چارچوب قدرتمند و قابل اعتماد برای پیش‌بینی دقیق و انعطاف‌پذیر تبدیل کرده است.

۳- پژوهش‌های پیشین

پژوهشی که توسط Mamykina و همکاران [۱(۲۰۱۱)] انجام شده است، عوامل موفقیت سایت

---

[۱] Weighted Stacking

[۲] Overfitting

[۳] Cross-Validation

پرسش و پاسخ **Stack Overflow** را بررسی می‌کند. این مطالعه نشان داد که موفقیت این پلتفرم به طراحی هوشمندانه، تعامل فعال بنیان‌گذاران با جامعه کاربری، و یک سیستم تشویقی مبتنی بر رفتارهای مطلوب بستگی دارد. نتایج تحقیق نشان می‌دهد که سایت دارای نرخ پاسخ‌دهی بیش از ۹۰ درصد و زمان میانگین ۱۱ دقیقه برای دریافت اولین پاسخ است. پژوهش همچنین به اکولوژی منحصربه‌فرد رفتار کاربران در این پلتفرم اشاره دارد که شامل گروه‌های مختلفی مانند فعالان جامعه، کاربران پرکار مقطعی، کاربران کم‌فعال و بازدیدکنندگان است. اگرچه این طراحی منجر به ایجاد یک منبع پایدار برای توسعه‌دهندگان نرم‌افزار شده است، محدودیت‌هایی نظیر مشکل «سریع‌ترین پاسخ‌دهنده» و تفاوت‌های میان کاربران تازه‌کار و خبره نیز شناسایی شده است. این پژوهش بر اهمیت طراحی پاسخ‌گو، به‌روزرسانی مداوم، و تعامل مستمر با کاربران برای موفقیت پلتفرم‌های اجتماعی تأکید می‌کند.

پژوهشی توسط Gkotsis و همکاران (۲۰۱۵) [۲]سیستم نوآورانه‌ای به نام **ACQUA** را برای پیش‌بینی بهترین پاسخ در وب‌سایت‌های پرسش و پاسخ معرفی می‌کند. این سیستم از ویژگی‌های متنی ساده برای ارزیابی پاسخ‌ها استفاده کرده و توانسته است عملکردی بهتر از روش‌های مبتنی بر زبان‌شناسی و معادل با روش‌های مبتنی بر امتیازدهی کاربران ارائه دهد. نتایج نشان می‌دهد که دقت میانگین سیستم ۸۴٪ و نرخ بازیابی آن ۷۰٪ است. این مطالعه همچنین نشان می‌دهد که گسسته‌سازی ویژگی‌های زبانی باعث بهبود چشمگیر عملکرد می‌شود، در حالی که ویژگی‌هایی مانند شهرت کاربران تأثیر قابل‌توجهی بر پیش‌بینی بهترین پاسخ ندارند. این تحقیق رویکردی موثر برای شناسایی پاسخ‌های مناسب در جوامع آنلاین ارائه کرده و بر کاربرد ویژگی‌های ساده متنی برای ارزیابی محتوای تولید شده توسط کاربران تأکید دارد.

در پژوهش Rajpurkar و همکاران (۲۰۱۶)]۳[، مجموعه‌داده Stanford Question Answering Dataset (SQuAD) معرفی شده است که شامل بیش از ۱۰۰,۰۰۰ پرسش و پاسخ از مقالات ویکی‌پدیا است. این مجموعه به عنوان چالشی برای مدل‌های درک مطلب طراحی شده است و تنوع بالایی از پرسش‌ها و انواع پاسخ‌ها را شامل می‌شود. پژوهش نشان داد که عملکرد انسانی (۸۶.۸٪) به مراتب بالاتر از بهترین مدل (۵۱٪) است، به‌ویژه با افزایش پیچیدگی پاسخ‌ها و تفاوت‌های نحوی میان سوال و جمله حاوی پاسخ. اگرچه مدل رگرسیون لجستیک برای برخی از انواع پاسخ‌ها عملکرد قابل قبولی داشت، اما در تشخیص دقیق محدوده پاسخ‌ها ضعف نشان داد. محدودیت‌های این پژوهش شامل تمرکز صرف بر مقالات ویکی‌پدیا و عدم توانایی مدل‌ها در مدیریت کامل تفاوت‌های نحوی است. بهبود مدل‌های پردازش زبان و گسترش مجموعه‌داده به موضوعات متنوع‌تر از پیشنهادهای آتی پژوهش است.

در سال‌های اخیر، توسعه سیستم‌های پرسش و پاسخ (QA)مبتنی بر پایگاه‌های دانش مورد توجه بسیاری از پژوهش‌ها قرار گرفته است. برای مثال، Cui و همکاران (۲۰۱۹)]۴[سیستم KBQA را معرفی کردند که از الگوهای یادگیری برای تبدیل سوالات طبیعی به پرسش‌های ساختاریافته استفاده می‌کند. این سیستم توانست با گسترش روابط در پایگاه دانش، پوشش پایگاه را تا ۵۷ برابر افزایش دهد و دقت بالایی در پاسخ به سوالات پیچیده به دست آورد. KBQA در مقایسه با روش‌های پیشرفته دیگر، در معیار QALD، دقت و بازخوانی بیشتری داشت و در شناسایی موجودیت‌ها نیز دقت ۷۲٪ را ثبت کرد. با این حال، عملکرد آن در پاسخ به سوالات غیردوتایی محدود بود که باعث کاهش بازخوانی و دقت کلی در برخی موارد شد.

در سال‌های اخیر، پژوهش‌های متعددی به بهبود عملکرد سامانه‌های پاسخ‌دهی به سوالات در حوزه‌های باز

پرداخته‌اند. Zhou و همکاران (۲۰۲۰) [۵] روش پاسخ‌دهی به سوالات حوزه باز مبتنی بر دانش (KAQA)را معرفی کردند که با استفاده از سه‌گانه‌های دانشی استخراج‌شده از منابع خارجی، روابط میان سوالات و اسناد و همچنین روابط میان اسناد را مدل‌سازی می‌کند. این چارچوب شامل سه مؤلفه اصلی است: بازیاب، خواننده و رتبه‌بند مجدد. آزمایش‌ها نشان داد که KAQA دقت بازیابی و رتبه‌بندی پاسخ‌ها را در مجموعه داده‌هایی نظیر SQuAD-open، Quasar-T وTriviaQA بهبود داده و عملکرد کلی سیستم‌های پاسخ‌دهی را افزایش می‌دهد. با این حال، این روش محدود به استفاده از مجموعه داده‌های مشخص بوده و از مدل‌های پیچیده‌تر مانند شبکه‌های گراف کانولوشنی برای گسترش دانش رابطه‌ای بهره نمی‌برد. این روش از بازیاب و رتبه‌بند مجدد مبتنی بر گراف‌های سوال-سند و سند-سند استفاده می‌کند که باعث بهبود دقت بازیابی و رتبه‌بندی پاسخ‌ها شده است. هرچند میزان دقت دقیق (Exact Match یا F۱) گزارش نشده است، اما تاکید پژوهشگران بر بهبود عملکرد کلی نسبت به سایر روش‌های موجود نشان‌دهنده تاثیر قابل‌توجه آن بر دقت سیستم‌های پاسخ‌دهی حوزه باز است.

در پژوهش Correa و (۲۰۱٤) Sureka [٦]، رفتار و ویژگی‌های سوالات حذف‌شده در Stack Overflow بررسی شده و چارچوبی پیش‌بینی‌کننده برای شناسایی سوالات باکیفیت پایین هنگام ایجاد پیشنهاد شده است. این مطالعه با تحلیل داده‌های پنج‌ساله و استفاده از ۴۷ ویژگی از جمله نمایه کاربر، محتوای سوال و سبک نحوی، مدلی با دقت ٪۶۶ در پیش‌بینی سوالات حذف‌شده ارائه کرده است. نتایج نشان داد که حدود ٪۸ از سوالات حذف می‌شوند، که اغلب توسط مدیران و پس از زمان قابل‌توجهی از دریافت اولین رأی حذف صورت می‌گیرد. سوالات حذف‌شده عموماً کیفیت بسیار پایینی دارند و در پایین ساختار هرمی کیفیت سوالات قرار می‌گیرند. محدودیت‌ها شامل عدم بررسی انواع دیگر محتوای بی‌کیفیت و نیاز به بهبود دقت مدل پیش‌بینی است. این پژوهش کاربردهایی در بهبود کیفیت محتوای پلتفرم‌های مشابه از طریق بازخورد فوری به کاربران و حمایت از مدیران دارد.

در پژوهش Jeon و همکاران (۲۰۰۶)]۷[، چارچوبی برای پیش‌بینی کیفیت پاسخ‌ها با استفاده از ویژگی‌های غیرمتنی توسعه داده شد. این مطالعه از روش‌هایی مانند "تخمین چگالی هسته‌ای" و "مدل آنتروپی بیشینه" برای پردازش داده‌ها استفاده کرده و نشان داده است که ترکیب این ویژگی‌ها در مدل‌های بازیابی اطلاعات می‌تواند عملکرد بازیابی پرسش‌های متداول را به طور قابل توجهی بهبود بخشد. آزمایش‌ها نشان داد که پیش‌بینی‌کننده توسعه‌یافته توانست پاسخ‌های باکیفیت را با دقت میانگین **۰.۹۲۲۷** و پاسخ‌های بی‌کیفیت را با دقت **۰.۶۵۵۸** از یکدیگر تفکیک کند. علاوه بر این، با تلفیق این معیار کیفیت در مدل‌های بازیابی، عملکرد رتبه‌بندی به طور معناداری بهبود یافت (با سطح اطمینان ۹۹٪). با این حال، محدودیت‌هایی مانند حجم کوچک داده‌های آموزشی (۸۹۴ نمونه آموزشی) و نیاز به توسعه معیارهای استاندارد برای سنجش کیفیت مستندات در پژوهش مشخص شد.

در پژوهش Shah و Pomerantz (۲۰۱۰)]۸[، کیفیت پاسخ‌ها در پلتفرم‌های پرسش و پاسخ اجتماعی مانند Yahoo! Answers ارزیابی و پیش‌بینی شد. این مطالعه دو رویکرد را برای سنجش کیفیت پاسخ‌ها مقایسه کرد: ارزیابی انسانی بر اساس ۱۳ معیار مشخص و استخراج خودکار ویژگی‌ها از داده‌ها. نتایج نشان داد که ویژگی‌های استخراج‌شده خودکار مانند نمایه کاربران و رتبه‌بندی متقابل پاسخ‌ها نسبت به معیارهای انسانی کارایی بهتری در پیش‌بینی پاسخ‌های باکیفیت دارند و دقت پیش‌بینی را از ٪۸۰.۳۳ به ٪۸۴.۱۷ افزایش دادند. با این حال، مدل لجستیک رگرسیون با استفاده از معیارهای انسانی نتوانست به‌طور موثر متغیرهای داده را توضیح دهد (با pseudo R^2برابر با ۰.۰۹۰۲). پژوهش همچنین بر

اهمیت اطلاعات اجتماعی و زمینه‌ای در ارزیابی کیفیت محتوا تأکید کرد. محدودیت‌های این مطالعه شامل کوچک بودن حجم داده‌ها و سوگیری احتمالی در ارزیابی انسانی است.

در سال‌های اخیر، بسیاری از پژوهش‌ها به بهبود شناسایی خودکار پست‌های کم‌کیفیت در پلتفرم‌های پرسش و پاسخ پرداخته‌اند. برای مثال، Ponzanelli و همکاران (۲۰۱۴) [۹] با استفاده از ویژگی‌های متنی و اجتماعی، مدلی مبتنی بر الگوریتم‌های ژنتیک ارائه دادند که توانست صف بررسی پست‌های کم‌کیفیت در Stack Overflowرا کاهش داده و پست‌های باکیفیت اشتباه‌شناسایی‌شده را با حداقل خطا حذف کند. این رویکرد با دستیابی به دقت نرم ۶۸.۹۵٪ و کاهش ۹ درصدی صف بررسی، زمان بررسی توسط مدیران را بهبود بخشید. با این حال، محدودیت داده‌های عمومی موجود از Stack Overflowممکن است بر دقت نتایج تأثیرگذار باشد. پیشنهاد شده است که این روش برای سیستم‌های توصیه‌گر ساده یا بهبود خودکار صف بررسی توسعه یابد.

همچنین، ارزیابی سیستم‌های پاسخ به پرسش توجه زیادی را به خود جلب کرده است. برای مثال، Chen و همکاران (۲۰۱۹) [۱۰] عملکرد معیارهای خودکار نظیر BLEU، ROUGE، METEOR و F۱ را بر روی سه مجموعه داده‌یNarrativeQA ، ROPES و SemEvalبررسی کردند. نتایج نشان داد که معیار METEORدر مجموعه داده‌هایNarrativeQA و SemEvalبیشترین تطابق را با ارزیابی‌های انسانی داشت، در حالی که F۱ در مجموعه داده ROPES عملکرد ضعیفی ارائه داد. همچنین، استفاده از معیار BERTScore که بازنمایی‌های متنی کانتکست‌محور را به کار می‌گیرد، بهبود جزئی نسبت به معیارهای سنتی نشان داد اما همچنان نیاز به بهینه‌سازی بیشتر دارد. این پژوهش به محدودیت معیارهای موجود در درک پاسخ‌های آزاد و پیچیده اشاره کرد و بر ضرورت توسعه ابزارهای دقیق‌تر برای ارزیابی سیستم‌های پاسخ به پرسش تأکید داشت.

Hu و همکاران (۲۰۱۷) [۱۱] رویکردی مبتنی بر یادگیری عمیق برای پیش‌بینی کیفیت پاسخ‌های خدمات پرسش و پاسخ آنلاین در حوزه سلامت پیشنهاد کردند که ویژگی‌های متنی و غیرمتنی را با استفاده از یک چارچوب شبکه باور عمیق چندرسانه‌ای[۴] ترکیب می‌کند. نتایج نشان داد که ویژگی‌های غیرمتنی، شامل ویژگی‌های زبانی سطحی و ویژگی‌های اجتماعی، نقش مؤثری در تمایز پاسخ‌های با کیفیت از پاسخ‌های کم‌کیفیت دارند. ویژگی‌های زبانی سطحی عملکرد بهتری نسبت به ویژگی‌های اجتماعی داشتند. این چارچوب نسبت به روش‌های پایه عملکرد بهتری نشان داد و دقت، فراخوانی و امتیاز F۱ بالاتری داشت. با این حال، مطالعه محدود به مجموعه داده خاصی از پلتفرمHaodf Online بود و ممکن است نتایج آن به سایر دامنه‌ها تعمیم‌پذیر نباشد.

Agichtein و همکاران (۲۰۰۸) [۱۲] چارچوبی برای ارزیابی کیفیت محتوای تولیدشده توسط کاربران در رسانه‌های اجتماعی، به‌ویژه در پورتال‌های پرسش و پاسخ جامعه‌ای مانند Yahoo! Answers، معرفی کردند. این رویکرد با ترکیب ویژگی‌های مبتنی بر محتوا، روابط کاربران و استفاده از داده‌ها (مانند بازدید صفحات)، به شناسایی محتوای با کیفیت بالا پرداخت. نتایج آزمایش‌ها

[۴] DBN

نشان داد که این سیستم می‌تواند محتوای با کیفیت را با دقت مشابه ارزیابی‌های انسانی شناسایی کند. در ارزیابی عملکرد، معیارهای دقت[۵]، فراخوانی[۶] و مساحت زیر منحنی(AUC) ROC گزارش شد که نشان‌دهنده عملکرد بالا و دقت مناسب مدل بود. مهم‌ترین ویژگی‌هایی که در این مطالعه شناسایی شدند شامل طول پاسخ‌ها، دسته‌بندی موضوعی و آمار استفاده از محتوا (مانند تعداد بازدید صفحات) بودند. همچنین، این مطالعه نشان داد که ویژگی‌های مختلف مانند روابط کاربران و آمار استفاده می‌توانند به دقت بیشتری در شناسایی محتوای با کیفیت کمک کنند. با این حال، محدودیت‌هایی وجود داشت؛ این روش به طور خاص برای پورتال‌های پرسش و پاسخ طراحی شده و ممکن است به سایر حوزه‌ها تعمیم‌پذیر نباشد. همچنین، شامل چالش‌هایی مانند بیش برازش در صورتی که تنها از آمار استفاده بدون در نظر گرفتن بازخوردهای دیگر استفاده شود، بود.

Blooma و همکاران (۲۰۰۸) [۱۳] در این مطالعه چارچوبی پیش‌بینی‌کننده برای ارزیابی کیفیت پاسخ‌ها در سیستم‌های پرسش و پاسخ مبتنی بر جامعه ارائه دادند. این چارچوب از ویژگی‌های متنی و غیرمتنی برای پیش‌بینی بهترین پاسخ استفاده می‌کند. نتایج نشان داد که ویژگی‌های متنی مانند دقت، تکمیل و منطقی بودن پاسخ، مهم‌ترین پیش‌بینی‌کننده‌ها برای کیفیت پاسخ بودند. ویژگی‌های غیرمتنی مانند شهرت و اعتبار پاسخ‌دهنده یا پرسش‌کننده تأثیر چندانی نداشتند. مطالعه با استفاده از یک تحلیل رگرسیون برای مدلسازی رابطه میان ویژگی‌ها و کیفیت پاسخ‌ها انجام شد. نتایج نشان داد که دقت، تکمیل و ارتباط پاسخ‌ها می‌توانند به‌عنوان پیش‌بینی‌کننده‌های کیفیت پاسخ استفاده شوند. مدل توسعه‌یافته دارای ضریب R برابر ۰.۸۸۶ بود که رابطه قوی بین مدل و کیفیت پاسخ را نشان می‌دهد. محدودیت‌های این مطالعه شامل استفاده از یک مجموعه داده خاص از Yahoo! Answers و محدودیت به یک دسته‌بندی خاص (علوم کامپیوتر) و نمونه‌گیری کوچک (۳۰۰ جفت سؤال و پاسخ) است. به علاوه، این مطالعه تنها بر ویژگی‌های متنی تمرکز داشت و ویژگی‌های غیرمتنی تأثیر کمی در کیفیت پاسخ داشتند. نتایج این مطالعه می‌توانند به بهبود ماژول‌های استخراج پاسخ در سیستم‌های پرسش و پاسخ و رتبه‌بندی و بازیابی پاسخ‌ها کمک کنند. تحقیقات آینده می‌توانند این یافته‌ها را برای سیستم‌های پرسش و پاسخ خودکار و گسترده‌تر استفاده کنند و تحلیل‌های مشابه را برای داده‌های بیشتری از سایر دسته‌ها انجام دهند.

در مطالعه‌ای که توسط Suggu و همکاران (۲۰۱۶) [۱٤] انجام شد، یک معماری یادگیری عمیق به نام DFFNبرای ارزیابی کیفیت پاسخ‌ها در سیستم‌های پرسش و پاسخ اجتماعی پیشنهاد شد. این معماری از ترکیب شبکه‌های عصبی کانولوشن و LSTM دوطرفه[۷] با توجه و ویژگی‌های دست‌ساز از منابع مختلف استفاده می‌کند. نتایج نشان داد که روش DFFN عملکردی برتر نسبت به روش‌های مرسوم دارد و در مجموعه داده‌های SemEval-۲۰۱۵ و SemEval-۲۰۱٦ نتایج بهتری را ارائه کرد. مدل پیشنهادی به‌ویژه DFFN-BLNA، بیشترین عملکرد را از خود نشان داد. این رویکرد با ترکیب ویژگی‌های استخراج‌شده از منابع خارجی مانند متن‌های ویکی‌پدیا و داده‌های کلیک، توانست ویژگی‌های تشابهی

[۵] Precision

[۶] Recall

[۷] BLSTM

را غنی‌تر کند. دقت مدل DFFN در مجموعه داده‌های SemEval-۲۰۱۵ و SemEval-۲۰۱۶ به طور خاص به صورت عددی ذکر نشده است، اما اشاره شده که این مدل عملکردی "state-of-the-art" داشت و توانست سیستم‌های پیشرفته‌ی دیگر را پشت سر بگذارد. بر اساس آزمایش‌های آماری انجام شده، نتایج به طور قابل توجهی بهتر از مدل‌های پایه بوده و در سطح اطمینان ۹۵٪ (α = ۰،۰۵) معنی‌دار بوده است. در نتیجه، مدل DFFN-BLNA بهترین عملکرد را در مقایسه با سایر مدل‌ها داشته است. در کل، نویسندگان ادعا کرده‌اند که این مدل به دقت بالاتر از سیستم‌های رقیب در همین مجموعه داده‌ها دست یافته است. این تحقیق همچنین محدودیتی در خصوص معایب یا کارهای آینده ارائه نکرده است. با وجود موفقیت‌های به‌دست‌آمده، یکی از چالش‌ها ممکن است نیاز به منابع داده‌ای غنی و زمان پردازش بالا باشد.

Hadfi و همکاران (۲۰۲۲) [۱۵] در مقاله خود به پیش‌بینی بهترین پاسخ در سایت‌های پرسش و پاسخ (CQA)با استفاده از اطلاعات کاربران پرداختند. در این مطالعه، ویژگی‌های مختلفی از جمله ویژگی‌های زبانی، اطلاعات کاربر، و ویژگی‌های رابطه‌ای برای پیش‌بینی بهترین پاسخ ترکیب شدند. نتایج نشان داد که روش پیشنهادی عملکرد بهتری نسبت به روش‌های قبلی داشته و ویژگی‌های زمان تاخیر بین سوال و پاسخ و امتیازدهی به پاسخ از توان پیش‌بینی بالاتری برخوردارند. در این تحقیق از مدل‌های مختلفی مانند SVM، Random Forest، MARS وLightGBM برای پیش‌بینی بهترین پاسخ استفاده شد که مدل LightGBM بهترین عملکرد را داشت. این روش به طور ویژه برای سایت‌های CQAمانند Stack Exchange مفید است، اما برای تعمیم یافته‌ها نیاز به ارزیابی‌های بیشتر با استفاده از داده‌های دیگر دارد. همچنین، محدودیت‌هایی مانند فرضیه جمعیت یکنواخت سوال‌کنندگان و حساسیت امتیازدهی به تغییرات پس از انتخاب بهترین پاسخ ذکر شده است.

Gkotsis و همکاران (۲۰۱۵) [۱۶] روشی نوین برای پیش‌بینی بهترین پاسخ در پلتفرم‌های پرسش و پاسخ مبتنی بر جامعه با استفاده از ویژگی‌های زبانی سطحی گسسته‌شده ارائه کردند. روش آن‌ها از روش‌های سنتی که به رتبه‌بندی و شهرت کاربران تکیه دارند، پیشی گرفت. این مطالعه نشان داد که استفاده از این ویژگی‌های گسسته‌شده، که شامل گروه‌بندی و مرتب‌سازی ویژگی‌های زبانی است، دقت بالاتری به همراه دارد و با F-Measure که یک معیار برای ارزیابی مدل‌های طبقه‌بندی است و ترکیبی از دقت و یادآوری است، برابر با ۰.۷۷ و AUC برابر با ۰.۸۸ عملکرد بهتری را از خود نشان می‌دهد. با وجود این که روش پیشنهادی عملکرد بهتری نسبت به روش‌های سنتی داشت، این مطالعه اشاره کرد که ویژگی‌های زبانی به تنهایی محدودیت‌هایی دارند، به ویژه زمانی که با استفاده از رتبه‌بندی کاربران مقایسه می‌شود. همچنین، نتایج ممکن است به دلیل تفاوت‌های زبانی در سایت‌های مختلفStackExchange متغیر باشد. پژوهش‌های آینده به بررسی قابلیت اعمال این روش در زمینه‌های دیگر و ارزیابی تأثیر کیفیت متنی در انتخاب پاسخ‌ها پیشنهاد شده است.

در سال‌های اخیر، پژوهش‌های مختلفی به ارزیابی کیفیت پاسخ‌ها در سایت‌های پرسش و پاسخ اجتماعی پرداخته‌اند. برای مثال، Zhu و همکاران (۲۰۰۹) [۱۷] مدلی چندبعدی برای ارزیابی کیفیت پاسخ‌ها معرفی کردند که شامل ۱۳ بعد مختلف مانند اطلاع‌رسانی، مودب بودن، تکمیل بودن، خوانایی، مرتبط بودن و صداقت است. این مدل از ترکیب تکنیک‌های کیفیت اطلاعات و پردازش زبان طبیعی استفاده می‌کند و قادر است فرآیند ارزیابی را به‌طور خودکار انجام دهد. نتایج اولیه نشان می‌دهد که این مدل می‌تواند دقت ۸۳.۹۸٪ را در پیش‌بینی پاسخ‌های خوب و بد بدست آورد. با این حال، برخی ابعاد مانند سطح جزئیات و تخصص نیاز به توسعه تکنیک‌های پردازش زبان طبیعی جدید دارند. این پژوهش‌ها به توسعه سیستم‌های ارزیابی کیفیت خودکار برای سایت‌های پرسش و پاسخ

اجتماعی کمک می‌کند، اما محدودیت‌هایی مانند استفاده از یک مجموعه داده خاص و نیاز به داده‌های بزرگتر برای بهبود دقت مدل وجود دارد.

Liu و همکاران (۲۰۱۰)]۱۸[مدلی برای پیش‌بینی بهترین پاسخ‌دهندگان در سایت‌های پرسش و پاسخ اجتماعی ارائه دادند. این مدل از ترکیب مدل زبان و تخصیص دیریکله نهفته[۸] برای مدل‌سازی پروفایل کاربران و پیش‌بینی پاسخ‌دهندگان استفاده می‌کند. با ترکیب فعالیت و اعتبار کاربران، مدل توانست دقت بالاتری در پیش‌بینی بهترین پاسخ‌دهندگان به دست آورد. نتایج تجربی با استفاده از معیار جدید "موفقیت در N[۹]" نشان داد که این روش می‌تواند به طور مؤثر سوالات جدید را به پاسخ‌دهندگان مناسب هدایت کند و کیفیت پاسخ‌ها را بهبود بخشد. با این حال، محدودیت‌هایی نظیر مقیاس کوچک داده‌ها و لزوم بهبود مدل‌سازی فعالیت و اعتبار کاربران وجود دارد. این روش کاربرد عملی در بهبود خدمات سایت‌های پرسش و پاسخ دارد، اما دقت پیش‌بینی می‌تواند به دلیل تعدد کاربران بالقوه کاهش یابد. دقت مدل به طور دقیق بیان نشده است، اما نتایج تجربی نشان داده‌اند که ترکیب مدل زبان و تخصیص دیریکله نهفته عملکرد بهتری نسبت به هر یک به تنهایی داشته است و افزودن اطلاعات مربوط به اعتبار کاربران و فعالیت کاربران نیز باعث بهبود دقت پیش‌بینی شده است. همچنین از معیاری به نام S@N[۱۰] برای ارزیابی عملکرد مدل استفاده شده است، که میزان موفقیت در توصیه بهترین پاسخ‌دهندگان را اندازه‌گیری می‌کند.

بلوم و همکاران (۲۰۱۰)]۱۹[در مطالعه‌ای به بررسی عوامل موثر بر انتخاب بهترین پاسخ در سرویس‌های پرسش و پاسخ جامعه‌محور[۱۱] پرداختند. آن‌ها از رویکردی ترکیبی برای تحلیل ویژگی‌های اجتماعی، متنی و ارزیابی محتوا استفاده کردند و داده‌های خود را از Yahoo! Answers جمع‌آوری نمودند. نتایج تحلیل رگرسیون لجستیک نشان داد که ویژگی‌های ارزیابی محتوا، نظیر کامل بودن و دقت، تاثیر بیشتری در انتخاب بهترین پاسخ دارند. همچنین، ویژگی‌های متنی مانند طول پاسخ و نسبت طول پرسش و پاسخ نیز مؤثر بودند، اما نقش کمتری نسبت به ویژگی‌های محتوا داشتند. از سوی دیگر، ویژگی‌های اجتماعی مانند تأیید کاربران ارتباط ضعیف‌تری با انتخاب بهترین پاسخ نشان دادند. این پژوهش، محدود به یک دسته خاص (کامپیوتر و اینترنت) از داده‌های Yahoo! Answers بوده و ممکن است به سایر دسته‌ها یا سرویس‌های مشابه تعمیم‌پذیر نباشد. دقت خاصی به‌عنوان یک معیار کمی گزارش نشده است. نتایج این تحقیق بیشتر بر تحلیل اهمیت نسبی ویژگی‌های مختلف (اجتماعی، متنی و ارزیابی محتوا) در انتخاب بهترین پاسخ تمرکز داشت. اگرچه آن‌ها به خوبی نشان داده‌اند که ویژگی‌های ارزیابی محتوا بیشترین تأثیر را در این فرآیند دارند، اما دقت مدل یا چارچوب پیشنهادی برای پیش‌بینی بهترین پاسخ در داده‌ها بیان نشده است. علاوه بر این، حجم داده‌های مورد استفاده کوچک بوده و نتایج بر اساس ارزیابی‌های کارشناسی استخراج شده است. از این رو، مطالعه‌های آتی می‌توانند با استفاده از داده‌های گسترده‌تر و دسته‌بندی‌های متنوع‌تر، نقش سایر عوامل را بررسی کنند. همچنین، پیشنهاد شده است از روش‌هایی نظیر تحلیل لینک برای

---

[۸] LDA

[۹] Success-at-N

[۱۰] Success-at-N

[۱۱] CQA

افزایش دقت و استفاده مجدد از پاسخ‌ها به‌عنوان منبع اطلاعاتی در سیستم‌های خودکار پرسش و پاسخ بهره گرفته شود.

Burel و همکاران (۲۰۱۲) [۲۰] مدلی برای شناسایی بهترین پاسخ‌ها در جوامع پرسش‌وپاسخ آنلاین طراحی کردند که از ویژگی‌های محتوا، کاربر و نخ‌های گفتگو بهره می‌برد. نتایج نشان داد ویژگی‌های نخ‌های گفتگو، مانند امتیازات و رتبه‌بندی‌ها، بیشترین تأثیر را در شناسایی بهترین پاسخ دارند، در حالی که ویژگی‌های کاربر، مانند نسبت شهرت موضوعی، نقش کمتری ایفا می‌کنند. ویژگی‌های نخ‌های گفتگو[۱۲] شامل اطلاعات ساختاری و تعاملی مربوط به یک موضوع در پلتفرم‌های پرسش‌وپاسخ است. این ویژگی‌ها شامل امتیازات و رأی‌ها (میزان رأی‌های مثبت و منفی)، تعداد پاسخ‌ها (کل پاسخ‌های ارسال‌شده در یک موضوع)، رتبه‌بندی پاسخ‌ها (مقایسه امتیاز یک پاسخ با دیگر پاسخ‌های همان موضوع)، زمان‌بندی‌ها (زمان ارسال پاسخ‌ها و ایجاد موضوع)، و ساختار نخ (ترتیب پاسخ‌ها و تعاملات کاربران) می‌شود. در مطالعه Burel و همکاران، این ویژگی‌ها تأثیر قابل‌توجهی در پیش‌بینی کیفیت پاسخ‌ها نشان داده و برتری آنها نسبت به ویژگی‌های مرتبط با محتوا و کاربران تأیید شده است. برخلاف پژوهش‌های قبلی، طول پاسخ با انتخاب بهترین پاسخ همبستگی معناداری نداشت. این مدل در سه مجموعه داده با دقت ۸۳٪ تا ۸۷٪ عملکرد موفقی داشت. با این حال، پژوهش محدود به سه جامعه آنلاین بوده و ممکن است به سایر پلتفرم‌ها قابل تعمیم نباشد.

Gkotsis و همکاران (۲۰۱۴) [۲۱] مدلی برای پیش‌بینی بهترین پاسخ در وب‌سایت‌های پرسش و پاسخ اجتماعی ارائه کردند که از ویژگی‌های زبانی سطحی گسسته‌شده استفاده می‌کند. این روش به دقت متوسط ۸۴٪ و بازخوانی ۷۰٪ دست یافت و عملکردی بهتر از روش‌های مبتنی بر زبان‌شناسی و مشابه با رویکردهای مبتنی بر امتیازدهی ارائه داد. در این پژوهش نشان داده شد که گسسته‌سازی ویژگی‌های زبانی میزان اطلاعات قابل استخراج را افزایش می‌دهد و استفاده از ویژگی‌های غیرزبانی تأثیری محدود بر عملکرد مدل دارد. این روش مستقل از داده‌های خاص هر جامعه بوده و بر روی ۲۱ وب‌سایت StackExchange آزمایش شده است. اگرچه مدل پیشنهادی برای شناسایی بهترین پاسخ مؤثر است، عدم استفاده از اطلاعات زمینه‌ای مانند امتیازات کاربران و پیچیدگی تغییرات زبانی در طول زمان محدودیت‌هایی ایجاد کرده است. پژوهش‌های آینده می‌توانند بر تحلیل تغییرات ویژگی‌های زبانی در طول زمان و استفاده از دیگر ویژگی‌ها برای بهبود عملکرد مدل تمرکز کنند.

Chen و همکاران (۲۰۱۲) [۲۲] روشی برای کالیبراسیون رأی‌ها در سیستم‌های پرسش و پاسخ اجتماعی پیشنهاد کردند که با کاهش اثرات سوگیری کاربران در رأی‌گیری، برآورد کیفیت پاسخ و تخصص کاربران را بهبود می‌بخشد. این روش از یک مدل یادگیری نظارت‌شده استفاده می‌کند که وزن‌های اهمیت رأی‌ها را محاسبه کرده و کیفیت پاسخ‌ها را پیش‌بینی می‌کند. نتایج نشان می‌دهد که مدل کالیبره‌شده در تخمین کیفیت پاسخ و رتبه‌بندی تخصص کاربران نسبت به روش‌های غیرکالیبره بهبود قابل‌توجهی در معیارهایی مانند NDCG@kو MAP دارد. با این حال، کوتاه بودن فهرست رتبه‌بندی در ارزیابی محدودیتی در تمایز عملکرد

[۱۲] Thread Features

روش‌های مختلف ایجاد کرده است. آینده پژوهش می‌تواند بر شناسایی عوامل دیگر مؤثر بر سوگیری رأی‌دهی و ادغام ویژگی‌های محتوایی در مدل‌ها تمرکز کند. روش پیشنهادی کاربردهای عملی در بهبود رتبه‌بندی پاسخ‌ها و کاربران متخصص در سیستم‌های پرسش و پاسخ اجتماعی دارد.

در پژوهشی توسط Elalfy و همکاران (۲۰۱۸) [۲۳]، یک مدل ترکیبی برای پیش‌بینی بهترین پاسخ در جوامع پرسش و پاسخ پیشنهاد شده است. این مدل شامل دو ماژول ویژگی‌های محتوایی و غیرمحتوایی است که ویژگی‌هایی همچون محتوای پاسخ، ارتباط میان پاسخ‌ها، و سطح اعتماد کاربران را تحلیل می‌کند. با استفاده از الگوریتم‌های طبقه‌بندی نظیر جنگل تصادفی و رگرسیون لجستیک، این پژوهش نشان داد که مدل ترکیبی در پیش‌بینی پاسخ‌های باکیفیت عملکرد بهتری نسبت به مدل‌های صرفاً محتوایی دارد. اگرچه این روش دقت بالایی را ارائه می‌دهد، محدودیت‌هایی همچون نیاز به داده‌های گسترده و هزینه پردازشی مطرح است. شکاف پژوهشی در این مطالعه، توجه ناکافی به نوع سوالات به‌عنوان ویژگی تأثیرگذار است که می‌تواند در پژوهش‌های آتی بررسی شود.

Zheng و Li (۲۰۱۷) [۲٤] روشی نوین به نام BoostStackرا برای پیش‌بینی بهترین پاسخ در Stack Overflowارائه کردند. این روش با بهره‌گیری از داده‌های ناهمگن مانند متن، کدهای نمونه و اطلاعات پس‌زمینه کاربران، از یادگیری چندنما استفاده می‌کند. الگوریتم BoostStack، با ترکیب Adaboost به‌عنوان طبقه‌بند سطح پایه و یک شبکه عصبی دو لایه به‌عنوان طبقه‌بند متا، دقت، بازخوانی و F-Measure بهتری نسبت به روش‌های تک‌نما ارائه می‌دهد. نتایج نشان می‌دهد که استفاده از اطلاعات چندنما عملکرد مدل را به‌طور قابل‌توجهی بهبود می‌بخشد، اما ترکیب صحیح نماها برای دستیابی به بهترین نتایج ضروری است. در حالی که الگوریتم برای Stack Overflow طراحی شده، قابلیت انتقال به سایر وب‌سایت‌های پرسش و پاسخ را نیز دارد. این پژوهش محدودیت خاصی را گزارش نکرده و می‌تواند در آینده با استفاده از الگوهای مشابه در سایر پلتفرم‌ها توسعه یابد. روش BoostStack در مقایسه با روش‌های تک‌نما دقت، بازخوانی و معیار F[۱۳] بالاتری را نشان داده است. به طور خاص، الگوریتم BoostStack در آزمایش‌ها بالاترین دقت و F-measure را به‌دست آورده و عملکرد بهتری نسبت به روش‌های پایه مانند Boost-NC، Boost-CCAP و Boost-CCAU داشته است. این نتایج تأیید می‌کنند که ترکیب داده‌های ناهمگن (متن، کد و اطلاعات پس‌زمینه) با استفاده از این روش، پیش‌بینی بهترین پاسخ‌ها را به طور چشمگیری بهبود می‌بخشد. دقت الگوریتم BoostStack برابر با ۷۷.۳٪ و معیارF برابر با ۷۶.۲٪ گزارش شده است. این مقادیر نسبت به سایر روش‌های مورد آزمایش، بالاتر بوده و نشان‌دهنده عملکرد برتر این الگوریتم در پیش‌بینی بهترین پاسخ‌ها در پلتفرم‌هایی مانندStack Overflow است.

توندوکار و همکاران (۲۰۱۸) [۲۵] از یک الگوریتم یادگیری رتبه‌بندی[۱۴] برای پیش‌بینی بهترین پاسخ‌دهنده به سوالات در انجمن‌های پرسش و پاسخ استفاده کردند. آن‌ها مجموعه جامعی از ویژگی‌ها شامل شباهت متنی، شباهت برچسب‌ها، و ویژگی‌های کاربرمحور نظیر تخصص

[۱۳] F-measure

[۱۴] LambdaMART

و در دسترس بودن را به کار بردند. مدل پیشنهادی در مقایسه با روش‌های موجود، عملکرد بهتری داشت و میانگین دقت رتبه‌بندی بالاتری[۱۵] ارائه داد. همچنین، مدل توانست بهترین پاسخ‌دهندگان را در ۰.۴۵ ثانیه به صورت بلادرنگ پیش‌بینی کند. از محدودیت‌های این پژوهش، عدم استفاده از داده‌های بزرگ‌تر و نبود امکان بازتولید نتایج مطالعات قبلی به دلیل دسترسی‌ناپذیری مجموعه داده‌ها بوده است.

چن و همکاران(Chen et al., ۲۰۱۷) [۲۶] سیستمی برای پاسخ به سوالات در دامنه باز[۱۶] با استفاده از ویکی‌پدیا به عنوان تنها منبع دانش[۱۷] ارائه کردند. این سیستم شامل یک بخش بازیابی اسناد[۱۸] مبتنی بر تطبیق دوگانه[۱۹] و TF-IDF و یک مدل شبکه عصبی چندلایه بازگشتی[۲۰] برای استخراج پاسخ‌ها از پاراگراف‌های ویکی‌پدیا است. مدل پیشنهادی در مجموعه داده SQuADبه نتایج پیشرفته‌ای دست یافت و از روش‌های موجود بهتر عمل کرد. استفاده از یادگیری چندوظیفه‌ای[۲۱] و نظارت دوردست [۲۲] عملکرد سیستم را بهبود بخشید.

با این حال، عملکرد در شرایط واقعی به دلیل چالش‌های خوانش ماشینی در مقیاس بزرگ[۲۳] محدود است. برای مثال، بازیابی اسناد تنها در ۷۷.۸٪ موارد پاسخ درست را بازیابی می‌کند. از محدودیت‌ها می‌توان به وابستگی به داده‌های ویکی‌پدیا و عدم امکان تعمیم به سایر حوزه‌ها اشاره کرد.

هو و همکاران (Hu et al., ۲۰۲۳) [۲۷] مدلی برای انتخاب پاسخ در سیستم‌های پرسش و پاسخ جامعه‌محور [۲۴]به نام شبکه توجه متقاطع پرسش-پاسخ[۲۵] یا QAN ارائه کردند که از مدل‌های پیش آموزش‌یافته[۲۶] و مدل‌های زبانی بزرگ[۲۷] بهره می‌برد. این مدل با استفاده از BERT برای رمزگذاری کلمات و مکانیزم توجه متقاطع [۲۸]، ویژگی‌های تعاملی بین سوالات و پاسخ‌ها را استخراج می‌کند. همچنین، از دانش خارجی تولیدشده توسط مدل‌های زبانی بزرگ برای بهبود دقت انتخاب پاسخ استفاده می‌شود.

مدل QAN بر روی دو مجموعه‌داده SemEval۲۰۱۵ و SemEval۲۰۱۷ به عملکرد پیشرفته‌ای دست یافت و در مقایسه با مدل‌های پایه در سه معیار ارزیابی (F۱، دقت Acc و میانگین دقت MAP) بهتر عمل کرد. بهینه‌سازی ورودی‌های اولیه[۲۹] برای LLM در ابعاد مختلف، مانند طول ورودی و موقعیت سوالات و پاسخ‌ها،

[۱۵] MRR@N
[۱۶] Open-Domain Question Answering
[۱۷] Knowledge Source
[۱۸] Document Retrieval
[۱۹] Bigram Matching
[۲۰] Multi-layer Recurrent Neural Network
[۲۱] Multitask Learning
[۲۲] Distant Supervision
[۲۳] Machine Reading at Scale
[۲۴] Community Question Answering
[۲۵] Question-Answer Cross Attention Network
[۲۶] Pre-trained Models
[۲۷] Large Language Models
[۲۸] Cross Attention Mechanism
[۲۹] Prompts

دقت انتخاب پاسخ‌ها را بهبود داده است.

با این حال، چالش‌های مربوط به تعامل جداگانه سوالات و پاسخ‌ها در روش‌های سنتی همچنان مطرح است. محدودیت‌های مدل به طور صریح بحث نشده‌اند، اما مطالعات نشان داده‌اند که هر بخش از مدل QAN برای دستیابی به عملکرد مطلوب ضروری است. بهبود سرعت محاسباتی و استفاده از مدل‌های پیش آموزش‌یافته جدید از اولویت‌های کارهای آینده معرفی‌شده است.

کاربردهای عملی این پژوهش شامل بهبود کارایی و دقت سیستم‌های CQA در حوزه‌های مختلف مانند خدمات مشتری، آموزش و بازیابی اطلاعات است.

توبا و همکاران (۲۰۱۴) [۲۸] چارچوبی سلسله‌مراتبی برای پیش‌بینی کیفیت پاسخ در سامانه‌های پرسش و پاسخ اجتماعی ارائه داده‌اند که از ترکیب طبقه‌بندی نوع پرسش و تحلیل کیفیت پاسخ استفاده می‌کند. این روش با تکیه بر ویژگی‌های ذاتی متن، از جمله تعداد کلمات بدون توقف[۳۰]، طول پاسخ و امتیاز خوانایی، دقت پیش‌بینی کیفیت پاسخ را بهبود می‌بخشد. نتایج نشان داد که این چارچوب در دسته‌بندی پاسخ‌های با کیفیت پایین، نسبت به روش‌های مبتنی بر یک طبقه‌بند، عملکرد بهتری دارد. این پژوهش از مدل یادگیری نظارت‌شده برای طبقه‌بندی نوع پرسش و یک اپراتور تصمیم‌گیری برای تحلیل کیفیت پاسخ استفاده می‌کند. مدل‌ها بر اساس ویژگی‌های معنایی، آماری، چگالی، و قطبیت احساسات آموزش داده شدند. همچنین سه روش تجمیع داده‌ها (SUM، MAX و STEP-۲) برای ارزیابی کیفیت پاسخ مقایسه شدند. چارچوب پیشنهادی حدود ۷۰ درصد دقت را روی مجموعه‌داده CQA به‌دست آورد و عملکرد بهتری نسبت به مدل‌های پایه ارائه داد. روش MAX بهترین نتایج را در پیش‌بینی کیفیت پاسخ‌ها نشان داد. همچنین، تحلیل‌ها نشان داد که کیفیت پاسخ‌ها می‌تواند بر اساس طول، ساختار، و محتوای آن‌ها سنجیده شود. مجموعه‌داده مورد استفاده به طور طبیعی نامتوازن بود (۸۰ درصد پاسخ‌های با کیفیت بالا)، که می‌تواند بر دقت چارچوب اثر بگذارد. همچنین، چالش‌هایی در تشخیص اسم‌ها در پاسخ‌ها وجود داشت که ممکن است فرآیند طبقه‌بندی را تحت تأثیر قرار دهد. پژوهشگران پیشنهاد می‌دهند که از ویژگی‌های اضافی، نظیر اطلاعات کاربر و دانش خارجی، برای بهبود دقت پیش‌بینی استفاده شود. همچنین، چارچوب پیشنهادی می‌تواند در سایر انواع محتواهای تولید شده توسط کاربران، نظیر پست‌های وبلاگ یا توییت‌ها، تعمیم یابد. این چارچوب قابلیت استفاده در بهبود کیفیت پاسخ‌ها در سامانه‌های پرسش و پاسخ نظیر Yahoo! Answers را دارد، که منجر به ارتقاء تجربه کاربر و کیفیت اشتراک‌گذاری دانش می‌شود.

کالفاتو و همکاران (Calefato et al., ۲۰۱۹) [۲۹] به ارزیابی عملکرد الگوریتم‌های مختلف طبقه‌بندی برای پیش‌بینی بهترین پاسخ در سایت Stack Overflow پرداختند. آن‌ها از معیارهای مستقل از آستانه مانند AUC و Balance استفاده کرده و ویژگی‌های کلیدی مؤثر بر دقت پیش‌بینی را شناسایی کردند. نتایج نشان داد که انتخاب الگوریتم طبقه‌بندی و تنظیم خودکار پارامترها تأثیر قابل‌توجهی بر عملکرد مدل‌ها دارد. در پیش‌بینی درون-پلتفرمی[۳۱]، مدل‌های برتر به عملکردی با میانگین AUC حدود ۰.۹ دست یافتند. با این حال، عملکرد

[۳۰] non-stop words

[۳۱] within-platform

مدل‌ها در پیش‌بینی بین-پلتفرمی[۳۲] با افتی ۲۰ تا ۲۷ درصدی همراه بود. همچنین تحلیل ویژگی‌ها نشان داد که ویژگی‌های مرتبط با امتیاز، قوی‌ترین پیش‌بین‌ها هستند. محدودیت‌های این پژوهش شامل تکیه بیش‌ازحد مدل‌ها به ویژگی‌های مبتنی بر امتیاز و عدم توجه به تأثیر زمان‌بندی در فرآیند پرسش و پاسخ است. این محدودیت‌ها ممکن است به کاهش عملکرد مدل‌ها در پیش‌بینی بین-پلتفرمی منجر شوند. در نتیجه، نویسندگان توصیه می‌کنند که تحقیقات آینده بر استفاده از ویژگی‌ها و تکنیک‌های جدید تمرکز کنند و از روش‌هایی مانند پردازش زبان طبیعی و تحلیل احساسات برای بهبود پیش‌بینی استفاده شود. این مطالعه کاربردهای عملی مهمی دارد، از جمله کمک به انتقال محتوا از فروم‌های قدیمی به پلتفرم‌های مدرن پرسش و پاسخ و تضمین کیفیت دانش جمع‌سپاری‌شده در این سایت‌ها. یافته‌ها می‌توانند برای طراحان این پلتفرم‌ها در ساخت ابزارهای خودکار انتخاب بهترین پاسخ مفید باشند.

در پژوهشی که توسط Gao و همکاران (۲۰۲۰) [۳۰] انجام شد، یک رویکرد مبتنی بر یادگیری عمیق به نام DeepAnsبرای شناسایی پاسخ‌های مرتبط‌تر در سایت‌های پرسش و پاسخ فنی معرفی گردید. این رویکرد به حل مشکل "تشنگی پاسخ[۳۳]" کمک می‌کند و نشان داده شده که عملکرد بهتری نسبت به مدل‌های مرسوم دارد. DeepAns از سه مرحله اصلی تشکیل شده است: تقویت پرسش[۳۴]، ایجاد برچسب[۳۵]، و توصیه پاسخ[۳۶]. در مرحله اول، پرسش‌های شفاف‌سازی با استفاده از مدل دنباله به دنباله[۳۷] ایجاد می‌شوند. در مرحله دوم، برچسب‌هایی برای داده‌ها شامل مثبت، خنثی+، خنثی- و منفی با استفاده از چهار قاعده تجربی ساخته می‌شود. در نهایت، مرحله توصیه پاسخ با استفاده از شبکه عصبی کانولوشنی[۳۸] امتیاز تطابق بین پرسش و پاسخ‌ها را محاسبه می‌کند. نتایج تجربی نشان می‌دهند که این رویکرد به طور چشمگیری از چندین مدل پایه موجود در ارزیابی خودکار پیشی می‌گیرد و در ارزیابی‌های انسانی نیز اثربخش بوده است. اگرچه این روش در حل مشکل تشنگی پاسخ موفقیت‌آمیز است، محدودیت‌های آن شامل تمرکز بر سؤالات مرتبط با زبان‌های برنامه‌نویسی پایتون و جاوا در سایت Stack Overflow به دلیل هزینه بالای مطالعه تجربی است. این رویکرد می‌تواند در سایت‌های پرسش و پاسخ فنی مانند Stack Overflow به توسعه‌دهندگان کمک کند تا پاسخ‌های مرتبط‌تر را شناسایی کنند و زمان جستجوی غیرضروری برای یافتن پاسخ‌های حل نشده را کاهش دهند.

در مطالعه Lou و همکاران (۲۰۲۳)]۳۱[، یک مرور جامع درباره‌ی پژوهش‌های مرتبط با پیروی از دستورات[۳۹] در پردازش زبان طبیعی[۴۰] ارائه شده است. این مطالعه موضوعاتی مانند تعریف دستورات وظیفه، مدل‌سازی دستورات، مجموعه داده‌های رایج و معیارهای ارزیابی و

[۳۲] cross-platform
[۳۳] answer hungriness
[۳۴] question boosting
[۳۵] label establishment
[۳۶] answer recommendation
[۳۷] sequence-to-sequence
[۳۸] CNN
[۳۹] Instruction Following
[۴۰] Natural Language Processing

چالش‌های پیش روی این حوزه را بررسی می‌کند. نویسندگان به اهمیت هم‌خوانی با اهداف پیش‌آموزش مدل‌های زبانی بزرگ و افزایش مقیاس مدل برای بهبود عملکرد اشاره کرده‌اند. همچنین، نتایج نشان می‌دهد که حتی مدل‌های پایه نیز می‌توانند عملکرد بهتری نسبت به مدل‌های کوچک‌تر تنظیم‌شده داشته باشند. محدودیت‌های این پژوهش شامل هزینه بالا و تأثیرات زیست‌محیطی مقیاس‌بندی مدل‌ها و ناکارآمدی روش‌های ارزیابی فعلی است. پیشنهاداتی برای آینده شامل توسعه روش‌های نمایشی جدید وظایف و رفع مشکلات مربوط به منفی‌سازی در دستورات ارائه شده است.

در مطالعه Sakai و همکاران (۲۰۱۱) [۳۲]، از معیارهای مرتبط با درجه‌بندی مرتبط بودن[۴۱] برای ارزیابی انتخاب پاسخ‌ها در سیستم‌های پرسش و پاسخ اجتماعی استفاده شده است. این پژوهش، داده‌های روزانه‌ی بیش از ۲۶۵۸ شرکت‌کننده با شرایط درد مزمن را در یک بازه زمانی ۱۵ ماهه تحلیل کرده و روابط معناداری بین درد و متغیرهای جوی مانند رطوبت نسبی، فشار و سرعت باد شناسایی کرده است. نتایج نشان داد که رطوبت نسبی بیشترین ارتباط را با درد دارد (احتمال وقوع رویداد درد با هر افزایش انحراف معیار در رطوبت نسبی ۱۲ درصد افزایش می‌یابد). روش‌های استفاده‌شده شامل طراحی مورد-مقابله‌ای[۴۲] و مدل‌های رگرسیون لجستیک شرطی بوده که تأثیر عواملی مانند خلق، فعالیت فیزیکی، و زمان سپری‌شده در خارج از منزل را نیز در تحلیل‌ها لحاظ کرده‌اند. از محدودیت‌های این پژوهش می‌توان به تعصب در انتخاب شرکت‌کنندگان، گزارش‌دهی ذهنی درد و عدم تعمیم‌پذیری به اقلیم‌های مختلف اشاره کرد. این پژوهش نشان می‌دهد که استفاده از داده‌های جمع‌آوری‌شده توسط ابزارهای شهروند-علمی[۴۳] می‌تواند درک ما از ارتباط بین متغیرهای محیطی و درد را بهبود بخشد و به توسعه پیش‌بینی‌های مرتبط با درد کمک کند.

والنتین (۲۰۲۲) [۳۳] از مدل‌های یادگیری عمیق شامل لایه‌های تعبیه[۴۴] و شبکه‌های حافظه طولانی‌مدت[۴۵] برای پیش‌بینی پاسخ‌های انتخاب‌شده توسط کاربران در Stack Overflow استفاده کرده است. این مدل‌ها به همراه ویژگی‌های عددی برای رتبه‌بندی پاسخ‌ها به کار گرفته شدند. هرچند استفاده از این روش‌ها باعث بهبود عملکرد نسبت به برخی مدل‌های پایه شد، اما نتوانستند عملکرد بهترین مدل‌های پایه را بهبود دهند. عملکرد مدل یادگیری عمیق در مقایسه با مدل‌های پایه به دلیل وجود تعداد زیادی کلمات خارج از واژگان[۴۶] یا OOV که ناشی از حضور کدهای برنامه‌نویسی در متن بود، محدود شد. همچنین، محدودیت طول توالی[۴۷] در مرحله تعبیه باعث از دست رفتن بخشی از اطلاعات شده است. نتایج نشان داد که مدل‌های پایه بر اساس ویژگی‌های عددی و متن، عملکرد بهتری نسبت به مدل‌های یادگیری عمیق داشتند. دقت دقیق مدل‌های یادگیری عمیق و پایه در مقاله به طور کمی ارائه نشده است، اما بیان شده که مدل یادگیری عمیق نتوانست به عملکرد بهترین مدل پایه که از

[۴۱] Graded-Relevance Metrics
[۴۲] Case-Crossover Design
[۴۳] Citizen-Science Experiments
[۴۴] Embedding
[۴۵] LSTM
[۴۶] Out-of-Vocabulary
[۴۷] Sequence Length

روش‌های کیسه کلمات[۴۸] استفاده می‌کرد، نزدیک شود. پیشنهاد شده که در تحقیقات آینده از مدل‌های پیشرفته‌تر مانند خانواده مدل‌های BERT استفاده شود و همچنین روش‌هایی برای شناسایی خودکار و حذف کدهای برنامه‌نویسی توسعه یابد تا مشکل OOV حل شود.

روی و همکاران (۲۰۲۳) [۳۴] در مقاله‌ای به بررسی سیستم‌های پرسش و پاسخ اجتماعی با استفاده از یادگیری ماشین و یادگیری عمیق پرداختند. آن‌ها پژوهش‌های موجود را در سه بخش اصلی دسته‌بندی کردند: پرسش‌ها، پاسخ‌ها و کاربران. این مطالعه نشان داد که اکثر تحقیقات بر کیفیت پرسش‌ها، کیفیت پاسخ‌ها و شناسایی کاربران خبره تمرکز دارند. در حالی که یادگیری ماشین در این حوزه غالب است، استفاده از یادگیری عمیق در حال افزایش است. مدل‌های ترکیبی و چندوجهی کمتر استفاده شده‌اند، و بسیاری از مدل‌ها تنها برای یک پلتفرم خاص آزمایش شده‌اند. از جمله نتایج، مدل‌های پیشنهادی به همبستگی پیرسون[۴۹] ۰,۶۲ و کوهن کاپا[۵۰] ۰,۸۶ دست یافتند. پژوهشگران نیاز به تحقیقات بیشتر در زمینه مدل‌های عمومی که قابلیت اجرا بر روی پلتفرم‌های مختلف را داشته باشند و استفاده از تکنیک‌های ترکیبی و چندوجهی را برجسته کردند.

روی و همکاران (۲۰۱۸) [۳۵] سیستمی را برای رتبه‌بندی پاسخ‌ها در سایت‌های پرسش و پاسخ اجتماعی بر اساس کیفیت آن‌ها پیشنهاد کردند. این سیستم، با استفاده از الگوریتم‌های یادگیری ماشین، ویژگی‌هایی شامل متنی و غیرمتنی (مانند تنوع لغوی، شهرت کاربر و شباهت پاسخ‌ها) را بررسی می‌کند و پاسخ‌های باکیفیت را در برگه‌ای به نام "پاسخ‌های امیدبخش[۵۱]" نمایش می‌دهد. مدل پیشنهادی با استفاده از ۲۶ ویژگی مختلف، از جمله تنوع اسمی و صفتی و شباهت بین پاسخ‌ها، دقت پیش‌بینی F۱-Score ۸۰% و بازخوانی ۹۰٪ را ارائه داده است. مدل Gradient Boosting در دسته‌بندی و مدل رگرسیون جنگل تصادفی[۵۲] در پیش‌بینی تعداد رأی‌ها عملکرد بهتری نسبت به مدل‌های بدون طبقه‌بند داشت. این سیستم، با میانگین دقت رتبه‌بندی[۵۳] برابر ۰.۸۶۵ و میانگین دقت متوسط[۵۴] برابر ۰.۴۸۳، توانایی پیش‌بینی بالای کیفیت پاسخ‌ها را نشان داد. محدودیت اصلی پژوهش استفاده از داده‌های محدود و تمرکز بر یک مجموعه داده خاص بود، اما پتانسیل تعمیم به سایر پلتفرم‌های CQA را نیز دارد.

زوراتو و همکاران (۲۰۲۳) [۳۶] تأثیر ویژگی‌های مختلف پاسخ‌ها را در پیش‌بینی بهترین پاسخ‌ها در انجمن‌های پرسش‌وپاسخ بررسی کرده‌اند. آن‌ها از الگوریتم‌های کلاسیک یادگیری ماشین مانند بیز ساده[۵۵]، رگرسیون لجستیک[۵۶] و جنگل تصادفی بر روی مجموعه داده‌ای شامل ۹۴۲۸ پست از Stack Overflow استفاده کردند. نتایج نشان داد که قابلیت خوانایی[۵۷] مهم‌ترین ویژگی برای تشخیص بهترین پاسخ‌ها است و به تنهایی

[۴۸] Bag of Words
[۴۹] Pearson correlation
[۵۰] Cohen kappa
[۵۱] Promising Answers
[۵۲] Random Forest Regression
[۵۳] MRR
[۵۴] MAP
[۵۵] Naïve Bayes
[۵۶] Logistic Regression
[۵۷] Readability

توانست دقت TPR برابر با ۰.۷۶۷ را به دست آورد. ترکیب قابلیت خوانایی و شباهت محتوایی[۵۸] نیز عملکرد قابل‌توجهی با TPR برابر ۰.۷۴۱ داشت. با این حال، نقش اجتماعی کاربران و تأثیر شهرت آنها به عنوان شکاف پژوهشی معرفی شده و نیازمند بررسی بیشتر است. این پژوهش نشان می‌دهد که بهبود خوانایی می‌تواند به ارتقاء کیفیت محتوای تولید شده و افزایش اعتبار کاربران کمک کند.

میمنو و همکاران (۲۰۱۱) [۳۷] در این پژوهش بهبود کیفیت موضوع‌ها در مدل‌های آماری موضوعی[۵۹] را از طریق معرفی یک معیار جدید برای ارزیابی انسجام موضوع‌ها بررسی کردند. آن‌ها از مدل آماری جایگزین *Polya urn* تعمیم‌یافته استفاده کردند که بهبود قابل‌توجهی در انسجام موضوع‌ها و کاهش موضوع‌های کم‌کیفیت نشان داد. این مدل بر اساس اطلاعات هم‌رخدادی واژگان طراحی شده و توانست در مجموعه‌ای از ۳۰۰,۰۰۰ خلاصه مقالات و پروژه‌های تحقیقاتی مؤسسات ملی بهداشت[۶۰] عملکرد بهتری نسبت به روش‌های سنتی مانند تخصیص دیریکله نهان ارائه دهد. نتایج نشان داد که این مدل انسجام متوسط موضوع‌ها را بهبود بخشیده و امتیاز انسجام برای ده موضوع با پایین‌ترین امتیاز را افزایش داده است. معیار انسجام پیشنهادی، موضوع‌های باکیفیت پایین را بهتر از معیارهایی مانند اندازه موضوع و اطلاعات متقابل نقطه‌ای[۶۱] شناسایی کرد. با این حال، مدل در کاهش کلی تعداد موضوع‌های نامناسب موفقیت چندانی نداشت و تنها ۱۳.۵٪ موضوع‌ها به‌عنوان "نامناسب" شناخته شدند که نسبت به ۱۶.۵٪ در LDA بهبود داشت اما از لحاظ آماری معنادار نبود ($p=۰,۰۵$). پژوهشگران همچنین به این نتیجه رسیدند که هیچ متدی قادر به پیش‌بینی کامل قضاوت‌های انسانی نیست. بهبود کیفیت معنایی موضوع‌ها، خصوصاً در سطح پایین و مقیاس‌پذیری برای مجموعه داده‌های بسیار بزرگ، به‌عنوان مسیرهای آتی تحقیق پیشنهاد شد. این پژوهش کاربردهای عملی در حوزه‌هایی مانند بازیابی اطلاعات و خلاصه‌سازی متون دارد و می‌تواند برای ایجاد مدل‌هایی با موضوعات خاص و باکیفیت در مجموعه‌های داده حوزه‌ای استفاده شود.

در مطالعه‌ای که توسط Molino و همکاران (۲۰۱۶) [۳۸] انجام شد، یک رویکرد جدید برای پیش‌بینی بهترین پاسخ در سایت‌های پرسش و پاسخ اجتماعی پیشنهاد شده است. این پژوهش از مجموعه ویژگی‌های گسترده شامل ویژگی‌های متنی، کاربری و شبکه‌ای استفاده کرده و بهبود ۱۱ تا ۲۶ درصدی در معیار دقت در رتبه یک[۶۲] نسبت به روش‌های پیشرفته پیشین نشان داده است. این تحقیق نشان داد که ویژگی‌های معنایی توزیعی[۶۳] می‌توانند جایگزین ویژگی‌های پرهزینه و زمان‌بر شباهت زبانی شوند و قدرت پیش‌بینی بالاتری ارائه دهند. همچنین، مشخص شد که کیفیت متن برای پیش‌بینی بهترین پاسخ در سوالات واقعی و ذهنی مناسب‌تر است، در حالی که ویژگی‌های مربوط به پروفایل کاربر برای سوالات بحث‌برانگیز و نظرسنجی‌ها کاربرد بیشتری دارند. محدودیت‌های این تحقیق مورد بررسی قرار نگرفته است،

[۵۸] Content Similarity
[۵۹] Topic Models
[۶۰] NIH
[۶۱] PMI
[۶۲] P@۱
[۶۳] Distributional Semantics

اما پیشنهاد شده که تحقیقات آینده به بررسی کارایی این ویژگی‌ها در سایر برنامه‌ها بپردازند.

انبارکی و جوکار (۲۰۲۱) [۳۹] به ارزیابی و پیش‌بینی کیفیت پاسخ‌ها در شبکه اجتماعی پژوهش‌محور[۶۴] در حوزه مدیریت دانش پرداختند. این مطالعه با تحلیل توصیفی و استنباطی بر روی ۵۴ سوال و ۴۴۳ پاسخ، که توسط ۳۰ متخصص از دو دانشگاه با استفاده از ۱۳ معیار ارزیابی شدند، انجام شد. نتایج نشان داد که متغیرهای ارتباط[۶۵]، کفایت[۶۶] و هماهنگی[۶۷] با ضرایب شانس به ترتیب ۳.۶۲۶، ۳.۴۴۰ و ۳.۱۴۸، بیشترین قدرت را در پیش‌بینی پاسخ‌های صحیح یا نادرست دارند. محدودیت‌های این پژوهش شامل حجم نمونه کوچک و ارزیابی ذهنی توسط متخصصین بود. پیشنهاد شده است که پژوهش‌های آینده بر توسعه روش‌های خودکار و عینی برای ارزیابی کیفیت پاسخ‌ها تمرکز کنند.

۳-۱- مقایسه با پژوهش‌های پیشین

پژوهش‌های پیشین در این حوزه از مدل‌های مختلفی مانندLogistic Regression ، Adaboost، SVM و شبکه‌های عصبی استفاده کرده‌اند. اما این پژوهش‌ها با محدودیت‌هایی از جمله دقت پایین، عدم توانایی ترکیب ویژگی‌های متنی و غیرمتنی و عدم استفاده بهینه از داده‌های محدود مواجه بوده‌اند. برخی از نتایج پژوهش‌های پیشین عبارت‌اند از:

- دقت ۶۶٪ با استفاده از Adaboost برای پیش‌بینی کیفیت سؤال.
- دقت ۸۴.۷۲٪ با Logistic Regression برای پیش‌بینی کیفیت پاسخ.
- دقت ۸۸٪ با مدل هیبریدی شامل Random Forest وLogistic Regression برای پیش‌بینی بهترین پاسخ.
- دقت ۹۵.۲٪ با Light Gradient Boosting Machineبرای پیش‌بینی بهترین پاسخ.

در مقایسه با این پژوهش‌ها، مدل پیشنهادی ما با دقت ۹۹.۰۳٪ نه تنها دقت بالاتری ارائه می‌دهد، بلکه از روش‌های پیشرفته‌ای مانند پشته‌ای وزن‌دار و ترکیب ویژگی‌های متنی و غیرمتنی بهره می‌برد. علاوه بر این، استفاده از مدل‌های پیشرفته‌ای مانند XLNet و RoBERTa، امکان استخراج معنای عمیق‌تر از متن و بهبود پیش‌بینی‌ها را فراهم کرده است.

این مدل در مقایسه با پژوهش‌های پیشین از دقت و قابلیت تعمیم‌دهی بالاتری برخوردار باشد. این دستاورد می‌تواند به عنوان یک الگوی جدید در توسعه سیستم‌های پرسش و پاسخ مورد استفاده قرار گیرد و تجربه کاربران در این حوزه را بهبود بخشد.

۳-۲- تفاوت و بهبودهای مدل ما نسبت به پژوهش‌های پیشین

در حالیکه بسیاری از پژوهش‌های پیشین در این حوزه مانند Burel و همکاران (۲۰۱۲) [۲۰] بر ویژگی‌های زبانی سطحی یا ویژگی‌های ساده متکی بودند، مدل ما از

---

[۶۴] ResearchGate

[۶۵] Relevance

[۶۶] Adequacy

[۶۷] Concordance

مدل‌های پیشرفته زبانی بهره گرفته که توانایی استخراج معنا و زمینه‌های پیچیده از متن را دارد. این امر باعث می‌شود عملکرد مدل در تحلیل دقیق‌تر پرسش‌ها و پاسخ‌ها بهبود یابد.

برخلاف روش‌هایی که به یک نوع ویژگی متکی بوده‌اند، مانند Zhu و همکاران (۲۰۰۹) [۱۷] (تمرکز بر ویژگی‌های متنی) یا Hadfi و همکاران(۲۰۲۲)]۱۵[(تمرکز بر ویژگی‌های غیرمتنی)، مدل ما ترکیبی از هر دو نوع ویژگی متنی و غیرمتنی را برای افزایش دقت پیش‌بینی استفاده کرده است.

مدل ما با ترکیب چندین یادگیرنده پایه قدرتمند مانند جنگل‌های تصادفی، XLNet و RoBERTa و به‌کارگیری یک فرایادگیر مبتنی بر درخت تصمیم، به دقت بالای ۹۹.۰۳٪ دست یافته است. این ترکیب، در مقایسه با مدل‌های منفرد مانند BoostStack (Zheng و Li، ۲۰۱۷)]۲۴[و یا DFFN (Suggu و همکاران, ۲۰۱۶)]۱۴[انعطاف‌پذیری و دقت بالاتری را تضمین می‌کند.

استفاده از وزن‌دهی مبتنی بر عملکرد هر مدل پایه، نوآوری دیگری است که در مطالعات پیشین کمتر دیده شده است. این تکنیک تأثیر مثبت بر کاهش خطای کلی مدل دارد.

با به‌کارگیری اعتبارسنجی متقاطع و ترکیب مدل‌های پایه، مدل ما توانسته است در شرایط کمبود داده‌های برچسب‌خورده نیز عملکرد مناسبی ارائه دهد، در حالی که برخی پژوهش‌ها مانند Elalfy و همکاران [۲۳] (۲۰۱۸)به محدودیت داده‌های گسترده اشاره کرده‌اند.

۳-۳- دقت و کارایی مدل ما نسبت به پژوهش‌های پیشین

دقت مدل ما بسیار بالاتر از مطالعات پیشین است. به‌عنوان مثال مدل BoostStack دقت ۷۷.۳٪ را گزارش داده است و DFFN عملکرد "state-of-the-art" داشته، اما دقت عددی ارائه نشده است. روش‌های ساده‌تر مانند ACQUA (Gkotsis و همکاران) دقت ۸۴٪ و نرخ بازیابی ۷۰٪ را ثبت کرده‌اند.

در حالی که بسیاری از مدل‌ها مانند روش پیشنهادی توسط Burel و همکاران (۲۰۱۲)]۲۰[محدود به داده‌های خاصی هستند، مدل ما با ترکیب ویژگی‌های متنوع توانایی تعمیم‌پذیری بیشتری دارد.

مدل ما با استفاده از ترکیب یادگیرنده‌ها و وزن‌دهی، از مشکلاتی مانند بیش‌برازش جلوگیری کرده است. این امر در مقایسه با مدل‌های پیچیده‌تر مانند DeepAns (Gao و همکاران)]۳۰[که به دلیل نیاز به هزینه پردازشی بالا محدود شده بودند، یک مزیت محسوب می‌شود.

۴-۳- شکاف‌های پژوهشی که مدل ما آن‌ها را پر کرده است

بسیاری از پژوهش‌ها بر یک نوع داده (متنی یا غیرمتنی) تمرکز داشته‌اند. مدل ما با ترکیب هر دو نوع داده، توانسته است شکاف موجود در تحلیل جامع کیفیت پاسخ‌ها را پر کند.

برخلاف مطالعاتی که از مدل‌های قدیمی‌تر یا ویژگی‌های ساده استفاده کرده‌اند (مانند ویژگی‌های زبانی سطحی در Gkotsis و همکاران، ۲۰۱۵]۱۶[)، مدل ما از جدیدترین تکنیک‌های پردازش زبان طبیعی بهره می‌برد.

با وجود پیچیدگی مدل‌های یادگیری عمیق، مدل ما با استفاده از فرایادگیر مبتنی بر درخت تصمیم و ترکیب هوشمندانه خروجی‌ها، قابلیت تفسیرپذیری را حفظ کرده است. همچنین، برخلاف پژوهش‌هایی که تنها در یک پلتفرم خاص آزمایش شده‌اند، مدل ما پتانسیل کاربرد در حوزه‌های مختلف را دارد.

استفاده از اعتبارسنجی متقاطع در مدل ما برای مقابله با کمبود داده‌های برچسب‌خورده، یک مزیت قابل توجه در

مقایسه با مدل‌هایی است که نیاز به داده‌های گسترده دارند، مانند روش Elalfy و همکاران (۲۰۱۸)]۲۳[.

٤- مدل پیشنهادی

مدل پیشنهادی این پژوهش، رویکردی ترکیبی و نظارت‌شده است که از قدرت یادگیری ژرف برای پردازش داده‌های پیچیده بهره‌برداری کرده و در عین حال، با انتخاب ویژگی‌های دستی برای داده‌های غیرمتنی، دقت و کارایی مدل را بهبود می‌بخشد. این ترکیب، که شامل استفاده از مدل‌های قدرتمند زبانی (مانند XLNet و RoBERTa) و روش‌های کلاسیک (مانند جنگل‌های تصادفی) است، توانسته است نیازهای متنوع و پیچیده جوامع پرسش و پاسخ آنلاین را به‌طور مؤثری برآورده کند.

مدل پینشهادی بر اساس دو اصل کلیدی طراحی شده است: الف) **استفاده از یادگیری ژرف برای داده‌های متنی:** مدل‌های یادگیری ژرف توانایی بالایی در استخراج خودکار ویژگی‌های پیچیده از داده‌های حجیم دارند، به‌ویژه در تحلیل داده‌های متنی که شامل ساختارهای نحوی و معنایی عمیق هستند. ب) **انتخاب ویژگی‌های دستی برای داده‌های غیرمتنی:** در مواردی که داده‌ها ساده‌تر و ساختارمندتر هستند (مانند آمار رأی‌گیری یا مشخصات کاربران)، استفاده از ویژگی‌های دستی می‌تواند ضمن کاهش پیچیدگی، تفسیرپذیری مدل را نیز افزایش دهد.

این مدل برای رفع شکاف‌های موجود در روش‌های پیشین و پاسخ به چالش‌های کلیدی این حوزه، نظیر تنوع در داده‌های متنی و غیرمتنی و نیاز به تفسیرپذیری نتایج، طراحی شده است. با ترکیب این دو رویکرد، مدل پیشنهادی توانسته است دقت پیش‌بینی را به ۹۹.۰۳٪ برساند و عملکردی فراتر از مدل‌های منفرد یا روش‌های ساده‌تر ارائه دهد.

٤-١- ساختار کلی مدل

مدل پیشنهادی در چندین فاز کلیدی طراحی و پیاده‌سازی شده است.

٤-١-١- پیش‌پردازش داده

این بخش شامل آماده‌سازی و تمیزسازی داده‌ها برای ورود به مدل می‌شود. در این پژوهش، هدف اصلی ارزیابی کیفیت پاسخ‌ها و انتخاب بهترین پاسخ در یک پلتفرم پرسش و پاسخ آنلاین است. برای تحقق این هدف، ابتدا داده‌های خام جمع‌آوری شده از پلتفرم با دقت مورد بررسی و پردازش قرار گرفتند. در مرحله اول، داده‌ها پاک‌سازی و ساختاردهی شدند تا برای استفاده در فرآیند مدل‌سازی آماده شوند. در ادامه، با تحلیل دقیق ویژگی‌های متنی و غیرمتنی مرتبط با سوالات و پاسخ‌ها، مجموعه‌ای از ستون‌های مؤثر و مرتبط برای مدل‌سازی انتخاب شد.

ویژگی‌های انتخاب‌شده شامل اطلاعات متنوعی نظیر متن سوالات و پاسخ‌ها، امتیازات کاربران، تعداد بازدیدها، تعاملات کاربران (مانند رأی‌ها و نظرات)، تغییرات محتوایی و زمانی، و شاخص‌های کیفیت پاسخ‌ها است. این ویژگی‌ها به‌گونه‌ای انتخاب شده‌اند که نه‌تنها نمایانگر رفتار کاربران و کیفیت محتوای تولیدشده باشند، بلکه امکان پیش‌بینی دقیق کیفیت پاسخ‌ها را نیز فراهم کنند.

در جدول ۱ ویژگی‌های استخراج‌شده از مجموعه‌داده نشان داده شده است.

جدول ۱: ویژگی‌های استخراج‌شده از مجموعه‌داده

نام ویژگی غیر متنی	نام ویژگی متنی

تعداد پاسخ‌ها	متن کامل سوال
امتیاز سوال	متن کامل پاسخ
تعداد بازدید برای جفت پست پرسش و پاسخ	
امتیاز پاسخ	
تعداد نظرات درج‌شده برای پاسخ	
رای به عنوان بهترین پاسخ توسط دیگر کاربران	
تعداد رای مثبت به پاسخ	
تعداد رای منفی به پاسخ	
تعداد رای به بسته شدن پاسخ	
تعداد رای به بازگشایی پاسخ	
تعداد رای به حذف پاسخ	
تعداد رای به بازگرداندن پاسخ حذف شده	

تغییر فاصله‌زمانی بین درج سوال تا درج پاسخ برحسب روز	
تعداد دفعات عنوان پاسخ	
تعداد دفعات تغییر محتوای پاسخ	
تعداد دفعات تغییر برچسب پاسخ	
پاسخ بسته‌شده	
پاسخ بازیابی شده	
درنظرگرفته‌شدن پاسخ به عنوان پاسخ جامعه	
پاسخ حذف‌شده	
میزان کیفیت پاسخ	

مجموعه‌داده نهایی پس از آماده‌سازی، به نسبت ۲۰/۸۰ برای آموزش و آزمون تقسیم شد. همچنین، داده‌های آموزشی نیز به نسبت ۲۰/۸۰ برای آموزش و اعتبارسنجی مدل تخصیص داده شدند. این تقسیم‌بندی تضمین می‌کند که مدل‌های یادگیری ماشین بتوانند به‌طور مؤثر آموزش ببینند و عملکرد آن‌ها به‌دقت ارزیابی شود.

این فرآیند پیش‌پردازش و انتخاب ویژگی‌ها، داده‌های اولیه را به مجموعه‌ای دقیق، تمیز و غنی تبدیل کرده که می‌تواند پایه‌ای مناسب برای توسعه مدل‌های پیش‌بینی کیفیت پاسخ‌ها باشد. این مرحله، نقشی حیاتی در بهبود عملکرد مدل‌ها و افزایش دقت پیش‌بینی‌ها ایفا می‌کند.

۲-۱-٤- ویژگی‌های متنی

در این پژوهش، از مدل‌های پیشرفته زبانی XLNet و RoBERTa برای تحلیل ویژگی‌های متنی پرسش‌ها و پاسخ‌ها در مجموعه‌داده StackOverflow استفاده شده است. این مدل‌ها که بر پایه معماری تبدیل‌کننده‌ها طراحی شده‌اند، قادر به درک روابط پیچیده میان کلمات و عبارات هستند و می‌توانند اطلاعات متنی را با دقت بسیار بالا پردازش و تحلیل کنند.

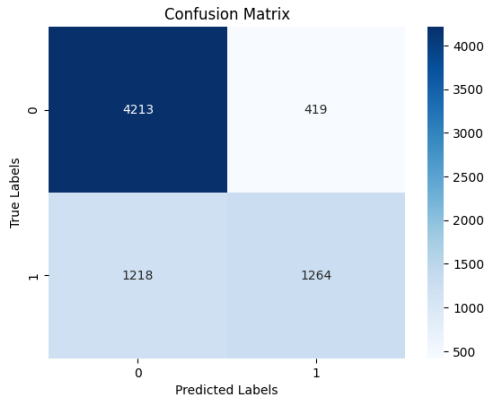
XLNet یکی از پیشرفته‌ترین مدل‌های زبانی است که با ترکیب مدل‌سازی زبان جایگشت و تکنیک‌های مدرن مانند Transformer-XL، محدودیت‌های مدل‌هایی مانند BERT را رفع کرده و عملکرد برتری در وظایف مختلف پردازش زبان طبیعی ارائه می‌دهد. این مدل با بهره‌گیری از احتمال شرطی توکن‌ها و ترتیب‌های مختلف، توانایی درک دقیق‌تر جملات و متون پیچیده را دارد.

جدول ۲ نشان می‌دهد که دقت نهایی[۶۸] مدل برابر با ۰.۷۷ است، که نشان می‌دهد مدل در ۷۷ درصد موارد به‌درستی بهترین پاسخ را پیش‌بینی کرده است. با این حال، تفاوت قابل‌توجهی بین معیارهای Precision و Recall برای کلاس‌های مختلف مشاهده می‌شود.

جدول ۲: گزارش طبقه‌بندی مدل XLNet برروی مجموعه‌داده آزمون

	دقت	بازخوانی	امتیاز F۱
۰	۰.۷۸	۰.۹۱	۰.۸۴
۱	۰.۷۵	۰.۵۱	۰.۶۱
دقت			۰.۷۷

ماتریس درهمی شکل ۱ نشان می‌دهد که مدل XLNet در تشخیص پاسخ‌های غیر برتر (کلاس ۰) عملکرد بهتری داشته است. از مجموع ۴۶۳۲ پاسخ غیر برتر، ۴۲۱۳ پاسخ به‌درستی تشخیص داده شده‌اند، اما ۱۲۱۸ پاسخ برتر به نادرست به عنوان پاسخ غیر برتر تشخیص داده شده‌اند. این نشان‌دهنده یک عدم تعادل در پیش‌بینی مدل است که باید در آینده با تنظیمات بیشتر بهبود یابد.



شکل ۱: ماتریس درهمی[۶۹] مدل XLNet برروی مجموعه‌داده آزمون

RoBERTa نسخه بهینه‌شده BERT است که با افزایش مقیاس داده‌های آموزشی و بهینه‌سازی تنظیمات، توانسته عملکرد خود را در وظایف پردازش زبان طبیعی بهبود

[۶۸] Accuracy

[۶۹] Confusion Matrix

بخشد. این مدل با تمرکز بر استخراج ویژگی‌های متنی دقیق‌تر، به تحلیل و تفسیر بهتر داده‌های متنی کمک می‌کند.

نتایج نشان داد که RoBERTa عملکرد بهتری نسبت به XLNet دارد و به دقت نهایی ۸۶.۹۴ درصد دست یافته است. این مدل توانایی بالایی در تشخیص پاسخ‌های برتر و غیر برتر داشته و توازن مناسبی میان معیارهای دقت، بازیابی و F۱ برقرار کرده است.

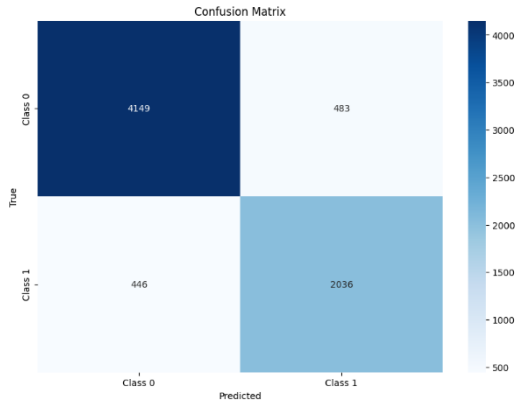
گزارش طبقه‌بندی نیز شامل معیارهایی نظیر دقت، فراخوانی، امتیاز F۱ و دقت کلی است. این گزارش به‌خوبی نقاط قوت و ضعف مدل در پیش‌بینی هر یک از کلاس‌ها را مشخص می‌کند. در جدول ۳ گزارش طبقه‌بندی مدل RoBERTa را برروی مجموعه‌داده آزمون مشاهده می‌کنیم.

جدول ۳: گزارش طبقه‌بندی مدل RoBERTa برروی مجموعه‌داده آزمون

	دقت	بازخوانی	امتیاز F۱
۰	۰.۹۰	۰.۹۰	۰.۹۰
۱	۰.۸۱	۰.۸۲	۰.۸۱
دقت			۰.۸۷

دقت نهایی مدل RoBERTa برروی مجموعه‌داده آزمون برابر با **۸۶.۹۴ درصد** به دست آمد که نشان‌دهنده کارایی مناسب این مدل در مقایسه با سایر روش‌های مورد استفاده در این پروژه بود.

در شکل ۲ ماتریس درهمی مدل RoBERTa را برروی مجموعه‌داده آزمون مشاهده می‌کنیم.



شکل ۲ - ماتریس درهمی مدل RoBERTa برروی مجموعه‌داده آزمون

هر دو مدل XLNet و RoBERTa توانستند اطلاعات متنی پرسش‌ها و پاسخ‌ها را با دقت بالایی تحلیل کنند. با این حال، RoBERTa با دقت ۸۶.۹۴ درصد عملکرد بهتری نسبت به XLNet (۷۷ درصد) داشت. همچنین، RoBERTa توانست توازن بهتری میان Precision و Recall برای هر دو کلاس برقرار کند.

استفاده از این مدل‌ها به‌عنوان بخشی از فرآیند کلی شناسایی بهترین پاسخ‌ها، نشان داد که انتخاب مدل مناسب و تنظیم دقیق آن می‌تواند تأثیر چشمگیری بر دقت پیش‌بینی‌ها و کارایی سیستم داشته باشد.

۳-۱-٤- ویژگی‌های غیرمتنی

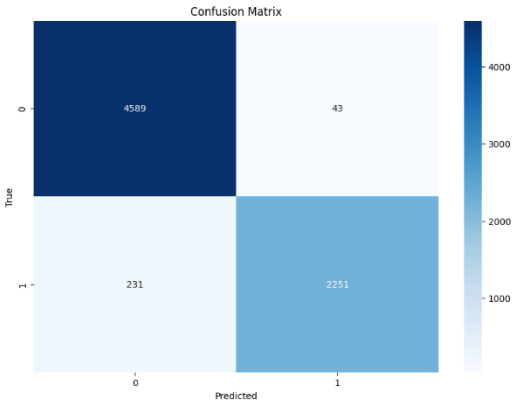
در این بخش، برای شناسایی بهترین پاسخ‌ها در سیستم‌های پرسش و پاسخ، ویژگی‌های غیرمتنی استخراج و مورد تحلیل قرار گرفتند. این ویژگی‌ها که شامل اطلاعات آماری و تعاملی هستند، به کمک مدل جنگل‌های تصادفی شناسایی و برای ورود به مدل‌های ترکیبی پردازش شدند. ویژگی‌های غیرمتنی شامل عواملی می‌شوند که از محتوای متن فراتر رفته و به جنبه‌های آماری، رفتاری و تعاملی کاربران و پاسخ‌ها می‌پردازند. به‌کارگیری این ویژگی‌ها در کنار ویژگی‌های متنی، به ایجاد مدلی جامع‌تر و با عملکرد بهتر منجر می‌شود. مدل جنگل‌های تصادفی که به‌صورت خاص برای تحلیل ویژگی‌های غیرمتنی طراحی شده است، به‌عنوان یکی از اجزای کلیدی مدل ترکیبی به کار گرفته شد. در ادامه،

مهم‌ترین ویژگی‌های غیرمتنی که در این پژوهش مورد استفاده قرار گرفتند در جدول ۴ نشان داده شده است.

جدول ۴: ویژگی‌های غیرمتنی مهم مورد استفاده قرار گرفته در مدل جنگل‌های تصادفی

ویژگی غیرمتنی	توضیح
امتیاز پاسخ	امتیازی که توسط کاربران به پاسخ داده شده است
امتیاز سوال	امتیازی که توسط کاربران به سوال داده شده است
تعداد بازدیدها	تعداد دفعاتی که سوال مشاهده شده است.
تعداد پاسخ‌ها	تعداد پاسخ‌هایی که به یک سوال داده شده است
تعداد نظرات پاسخ	تعداد نظراتی که برای یک پاسخ داده شده است
تاریخ ایجاد پاسخ	زمان ایجاد پاسخ و مدت زمانی که از انتشار آن گذشته است
فعالیت‌های امتیازی	مانند تعداد رای‌های مثبت و منفی که پاسخ دریافت کرده است و یا رأی به بسته شدن یا باز شدن و حتی رأی به حذف شدن یا نشدن
فعالیت‌های مدیریتی	مانند بسته شدن، باز شدن مجدد، حذف شدن و بازگردانی پاسخ و یا حتی تغییر متن و عنوان. همچنین تغییرات در برچسب‌ها نیز جز این عملیات به شمار می‌رود

پس از آموزش و ارزیابی مدل، نتایج نشان دادند که مدل جنگل‌های تصادفی توانایی بسیار بالایی در تشخیص ویژگی‌های غیرمتنی مرتبط با کیفیت پاسخ‌ها دارد. دقت مدل، که در جدول ۴ گزارش شده است، برای مجموعه داده آزمون به ۹۶ درصد رسید، که نشان‌دهنده عملکرد قوی مدل در شناسایی بهترین پاسخ‌ها است. مدل توانست ۴۵۸۹ نمونه غیر برتر را به درستی شناسایی کند، در حالی که تنها ۴۳ نمونه را به اشتباه به‌عنوان برتر طبقه‌بندی کرد. با این حال، ضعف مدل در کاهش نرخ False Negative همچنان قابل توجه است. مقادیر دقیق ماتریس درهمی در شکل ۳ نشان داده شده است.



شکل ۳: ماتریس درهمی مدل جنگل‌های تصادفی برروی مجموعه‌داده آزمون

جدول ۴: گزارش طبقه‌بندی مدل جنگل‌های تصادفی برروی مجموعه‌داده آزمون

	دقت	بازخوانی	امتیاز F۱
۰	۰.۹۵	۰.۹۹	۰.۹۷
۱	۰.۹۸	۰.۹۱	۰۹۴
دقت			۰.۹۶

این نتایج نشان می‌دهد که مدل جنگل‌های تصادفی، با استفاده از ویژگی‌های غیرمتنی، به‌خوبی قادر به تفکیک پاسخ‌های با کیفیت بالا و پایین است و نقش کلیدی در تقویت مدل ترکیبی دارد.

٤-١-٤- ترکیب مدل‌ها: به‌کارگیری رأی‌گیری اکثریت با مدل‌های پایه

در این بخش، تکنیک رأی‌گیری اکثریت به‌عنوان یک روش یادگیری جمعی برای ترکیب پیش‌بینی‌های مدل‌های مختلف به کار گرفته شده است. هدف اصلی این روش، بهبود عملکرد کلی و افزایش دقت پیش‌بینی‌ها از طریق ترکیب خروجی مدل‌های پایه است. مدل‌های به‌کاررفته شامل **XLNet** و **RoBERTa** برای تحلیل ویژگی‌های متنی و **جنگل‌های تصادفی** برای تحلیل ویژگی‌های غیرمتنی هستند .در این پژوهش، از روش رأی‌گیری سخت استفاده شده است که بر اساس بیشترین تعداد آرا، کلاس نهایی را تعیین می‌کند.

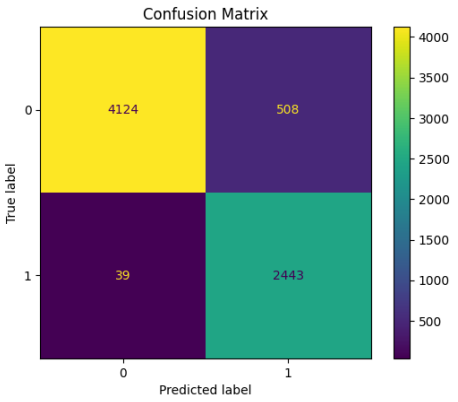
برای ارزیابی مدل رأی‌گیری اکثریت، پیش‌بینی‌های خروجی مدل‌های پایه **XLNet**، **RoBERTa** و **جنگل‌های تصادفی** بر روی داده‌های آموزشی، اعتبارسنجی و آزمون ترکیب شدند. این روش از نقاط قوت هر مدل بهره‌برداری کرده و نقاط ضعف آن‌ها را پوشش داده است. حال نتایج ارزیابی نشان می‌دهند که روش رأی‌گیری اکثریت توانسته است عملکرد قابل‌توجهی بر روی داده‌های آزمون ارائه دهد.

ماتریس درهمی شکل ۴ و گزارش طبقه‌بندی جدول ۵ بیانگر این موضوع است. ماتریس درهمی شکل ۴ نشان می‌دهد که مدل توانسته است تعداد ۴۱۲۴ نمونه از کلاس ۰ و ۲۴۴۳ نمونه از کلاس ۱ را به‌درستی دسته‌بندی کند. بااین‌حال، تعداد ۵۰۸ نمونه از کلاس ۰ به نادرست به کلاس ۱ و ۳۹ نمونه از کلاس ۱ به نادرست به کلاس ۰ طبقه‌بندی شده‌اند.

جدول ۵ نشان می‌دهد که دقت کلی مدل برابر با **٪۹۲.۳۱** است. همچنین، معیارهای دقت، بازخوانی و امتیاز F۱ برای کلاس‌های ۰ و ۱ محاسبه شده‌اند.

جدول ۵: گزارش طبقه‌بندی مدل رأی‌گیری اکثریت با مدل‌های پایه XLNet، جنگل‌های تصادفی و RoBERTa برروی مجموعه‌داده آزمون

	دقت	بازخوانی	امتیاز F۱
۰	۰.۹۹	۰.۸۹	۰.۹۴
۱	۰.۸۳	۰.۹۸	۰.۹۰
دقت			۰.۹۲

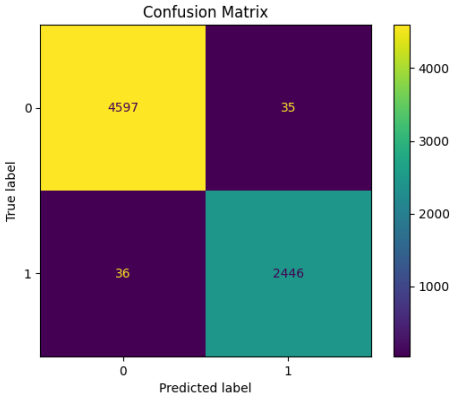


شکل ۴: ماتریس درهمی مدل رأی‌گیری اکثریت با مدل‌های پایه XLNet، جنگل‌های تصادفی و RoBERTa برروی مجموعه‌داده آزمون

مدل رأی‌گیری اکثریت، با ترکیب پیش‌بینی‌های سه مدل پایه XLNet، RoBERTa و **جنگل‌های تصادفی**، توانسته است بهبود چشمگیری در دقت و تعمیم‌پذیری پیش‌بینی‌ها ایجاد کند. این روش با ترکیب قابلیت‌های متمایز هر مدل، عملکردی قوی و پایدار ارائه داده است.

۵-۱-٤- فرایادگیر: به‌کارگیری مدل یادگیری جمعی پشته‌ای وزن‌دار با مدل‌های پایه و فرایادگیر درخت تصمیم

در این بخش، از تکنیک یادگیری جمعی پشته‌ای برای ترکیب مدل‌های مختلف به‌منظور بهبود دقت پیش‌بینی‌ها و مدیریت نتایج ترکیبی استفاده شده است. پشته‌ای یک روش پیشرفته در یادگیری ماشین است که در آن چندین مدل پایه برای یادگیری ویژگی‌های متنوع داده‌ها به کار گرفته می‌شوند. پیش‌بینی‌های این مدل‌ها به‌عنوان ورودی به یک مدل **فرایادگیر**[۷۰] داده می‌شوند تا پیش‌بینی نهایی انجام شود. در این پژوهش، مدل‌های پایه شامل XLNet و RoBERTa برای تحلیل و استخراج ویژگی‌های معنایی متنی با استفاده از یادگیری ژرف، جنگل‌های تصادفی، برای شناسایی و تحلیل ویژگی‌های غیرمتنی و رأی‌گیری اکثریت، به‌عنوان یک ویژگی جدید، ترکیب پیش‌بینی‌های مدل‌های پایه، می‌شود. همچنین از درخت تصمیم به‌عنوان مدل فرایادگیر استفاده شد چراکه درخت تصمیم با توانایی پردازش ترکیب پیچیده ویژگی‌ها را دارد. با توجه به جدول ۶، مدل پشته‌ای وزن‌دار به دقت ٪۹۹ بر روی داده‌های آزمون دست یافت که نشان‌دهنده پیش‌بینی‌های بسیار دقیق و قابل‌اعتماد است. معیارهای دقت، بازخوانی و امتیاز F۱ برای هر کلاس در جدول ۶ ارائه شده‌اند و بیانگر عملکرد فوق‌العاده مدل در شناسایی صحیح نمونه‌ها هستند. در شکل ۵، تعداد بسیار کمی از نمونه‌ها به‌نادرست پیش‌بینی شده‌اند که این امر نشان‌دهنده کاهش خطاهای مدل است.



شکل ۵: ماتریس درهمی مدل یادگیری جمعی پشته‌ای وزن‌دار با مدل‌های پایه XLNet، جنگل‌های تصادفی، RoBERTa و رأی‌گیری اکثریت و مدل فرایادگیر درخت تصمیم برروی مجموعه‌داده آزمون

جدول ۶: گزارش طبقه‌بندی مدل یادگیری جمعی پشته‌ای وزن‌دار با مدل‌های پایه XLNet، جنگل‌های تصادفی، RoBERTa و رأی‌گیری اکثریت و مدل فرایادگیر درخت تصمیم برروی مجموعه‌داده آزمون

	دقت	بازخوانی	امتیاز F۱
۰	۰.۹۹	۰.۹۹	۰.۹۹
۱	۰.۹۹	۰.۹۹	۰.۹۹

[۷۰] Meta-Learner

۰.۹۹			دقت

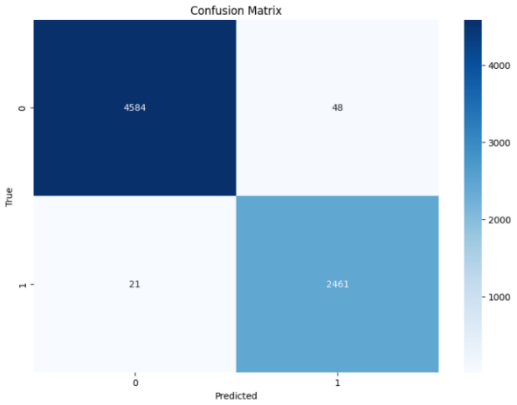
روش پشته‌ای با ترکیب مدل‌های قدرتمند XLNet، RoBERTa، جنگل‌های تصادفی و رأی‌گیری اکثریت توانست نقاط قوت هر مدل را بهینه کرده و نقاط ضعف آن‌ها را پوشش دهد. استفاده از **درخت تصمیم** به‌عنوان مدل فرایادگیر موجب بهبود قابل‌توجه دقت پیش‌بینی‌ها شد.

این مدل، با دستیابی به دقت ۹۹٪، نشان داد که ترکیب مناسب مدل‌های پایه با روش‌های یادگیری جمعی می‌تواند بهبود چشمگیری در مسائل دسته‌بندی ایجاد کند و به‌عنوان یک چارچوب قابل‌اعتماد برای حل مسائل مشابه به کار گرفته شود.

۲-٤- بهبود مدل یادگیری جمعی پشته‌ای وزن‌دار با استفاده از درصد وقوع کلاس‌ها

برای رفع مشکلات موجود در مدل یادگیری جمعی پشته‌ای وزن‌دار، شامل مدل‌های پایه XLNet، RoBERTa، جنگل‌های تصادفی و رأی‌گیری اکثریت، رویکردی جدید مبتنی بر استفاده از درصد وقوع کلاس‌ها به‌جای پیش‌بینی دقیق کلاس در مدل فرایادگیر درخت تصمیم ارائه شد. این مشکلات شامل انتخاب بیش از یک پاسخ به‌عنوان بهترین پاسخ یا عدم انتخاب هیچ پاسخی برای برخی سوالات بود. به‌جای دریافت خروجی‌های صفر و یک از مدل فرایادگیر، خروجی قبل از لایه Softmax که نشان‌دهنده احتمال وقوع هر کلاس است، استفاده شد. این روش به مدل اجازه می‌دهد برای هر پاسخ، درصد تعلق آن به کلاس ۱ (بهترین پاسخ) یا کلاس ۰ (عدم بهترین پاسخ) را محاسبه کند. سپس پاسخ‌ها برای هر سوال بر اساس درصد وقوع مرتب شده و پاسخی با بالاترین احتمال وقوع به‌عنوان بهترین پاسخ انتخاب می‌شود. این اقدام تضمین می‌کند که برای هر سوال، دقیقاً یک پاسخ به‌عنوان بهترین پاسخ انتخاب شود و از مشکلات پیشین جلوگیری شود.

پس از اعمال این رویکرد، دقت و کارایی مدل به طور چشمگیری بهبود یافت. خروجی نهایی مدل، شامل شاخص‌های عملکرد و ماتریس درهمی، نشان داد که مدل توانسته است با دقت **۹۹.۰۳٪** عملکردی بسیار قوی ارائه دهد. نتایج دقیق در جدول ۷ و ماتریس درهمی شکل ۶ به‌طور کامل ارائه شده است.



شکل ۶: ماتریس درهمی مدل یادگیری جمعی پشته‌ای وزن‌دار با مدل‌های پایه XLNet، جنگل‌های تصادفی، RoBERTa و رأی‌گیری اکثریت و مدل فرایادگیر درخت تصمیم برروی مجموعه‌داده آزمون

جدول ۷: گزارش طبقه‌بندی مدل یادگیری جمعی پشته‌ای وزن‌دار با مدل‌های پایه XLNet، جنگل‌های تصادفی، RoBERTa و رأی‌گیری اکثریت و مدل فرایادگیر درخت تصمیم و استفاده از خروجی درجه وقوع هر کلاس به جای پیش‌بینی دقیق کلاس در مدل فرایادگیر برروی مجموعه‌داده آزمون

	دقت	بازخوانی	امتیاز F۱
۰	۰.۹۹	۱.۰۰	۰.۹۹

۰.۹۹	۰.۹۸	۰.۹۹	۱
۰.۹۹			دقت

مدل بهبود یافته با استفاده از درصد وقوع کلاس‌ها توانست مشکلات پیشین را برطرف کرده و دقت و بازخوانی خود را به سطحی بی‌نظیر برساند. با وجود تعداد کمی نمونه که به اشتباه طبقه‌بندی شده‌اند، این روش به دلیل عملکرد قوی در شناسایی بهترین پاسخ‌ها، پایه‌ای مطمئن برای بهبودهای آینده فراهم می‌کند. رویکرد پیشنهادی، نه‌تنها دقت بالایی ارائه می‌دهد، بلکه روشی تعمیم‌پذیر برای مسائل مشابه در دیگر جوامع پرسش و پاسخ نیز محسوب می‌شود.

بلوک دیاگرام ارائه‌شده در شکل ۷ ساختار کلی مدل پیشنهادی را برای پیش‌بینی بهترین پاسخ در جوامع پرسش و پاسخ نشان می‌دهد. در این بلوک دیاگرام، داده‌های ورودی ابتدا به دو بخش متنی و غیرمتنی تفکیک می‌شوند. ویژگی‌های متنی و غیرمتنی به‌طور جداگانه پیش‌پردازش می‌شوند تا برای مراحل بعدی آماده شوند. این پیش‌پردازش شامل پاک‌سازی داده‌ها، نرمال‌سازی متون و استخراج ویژگی‌های خاص است. سپس داده‌های پیش‌پردازش‌شده به چهار مدل پایه تغذیه می‌شوند. مدل‌های XLNet و RoBERTa برای استخراج ویژگی‌های متنی با کیفیت بالا، **جنگل‌های تصادفی** برای تحلیل ویژگی‌های غیرمتنی، سایر مدل‌های ترکیبی برای تقویت دقت و افزایش پوشش خطاها مورد استفاده قرار گرفتند. سپس خروجی مدل‌های پایه از طریق یک بلوک ترکیب وزن‌دار به‌طور هوشمندانه وزن‌دهی می‌شود. این فرآیند باعث بهره‌برداری حداکثری از مزایای هر مدل و پوشش ضعف‌های آن‌ها می‌شود. خروجی‌های وزن‌دهی‌شده مدل‌ها در یک بلوک پشته‌ای، ادغام می‌شوند. این مرحله با استفاده از یک متا-یادگیرنده که بر اساس **درخت تصمیم‌گیری** طراحی شده است، پیش‌بینی نهایی را انجام می‌دهد. برای اطمینان از دقت مدل، پیش‌بینی نهایی بر اساس درصد وقوع خروجی‌ها انجام می‌شود و نه پیش‌بینی قطعی صفر و یک. این رویکرد انعطاف‌پذیری بیشتری را در تحلیل داده‌ها فراهم می‌کند. مدل از طریق اعتبارسنجی متقاطع مورد آزمایش قرار گرفته و عملکرد آن با دقت بالا بهبود یافته است.

این ساختار ترکیبی و هوشمندانه از قابلیت‌های مدل‌های پیشرفته زبانی، یادگیری ماشین و یادگیری جمعی بهره‌برداری می‌کند تا عملکردی بی‌نظیر در پیش‌بینی بهترین پاسخ‌ها ارائه دهد.

۵- مجموعه‌داده

برای انجام این پژوهش، از مجموعه‌داده‌های **StackExchange** استفاده شده است. این مجموعه‌داده یکی از منابع جامع و ساختاریافته برای تحلیل داده‌های مرتبط با جوامع پرسش و پاسخ است و شامل اطلاعات متنوعی از سایت‌های مختلف در زیرمجموعه StackExchange مانند **StackOverflow**، **Super User** و **Android Enthusiasts** می‌باشد. داده‌های این مجموعه به صورت فایل‌های **XML** ارائه شده‌اند و هر فایل حاوی اطلاعاتی درباره پست‌ها، کاربران، نظرات، برچسب‌ها و سایر تعاملات موجود در این پلتفرم‌ها است.

در این پژوهش، از زیرمجموعه **android.stackexchange.com** استفاده شده است که بر موضوعات مرتبط با اندروید تمرکز دارد. این مجموعه داده شامل پرسش‌ها، پاسخ‌ها و سایر اطلاعات تعاملی کاربران در حوزه توسعه اندروید است و برای تحلیل کیفیت پاسخ‌ها و شناسایی بهترین پاسخ‌ها بسیار مناسب است. داده‌ها به دلیل ساختار دقیق و ویژگی‌های متنوعی که ارائه می‌دهند، امکان استخراج ویژگی‌های متنی و غیرمتنی متعددی را برای مدل‌سازی فراهم می‌کنند. انتخاب این مجموعه به دلیل تمرکز آن بر حوزه اندروید و قابلیت تعمیم نتایج به سایر جوامع پرسش و پاسخ انجام شده است.

۶- جمع‌بندی

در این پژوهش، مدلی کارآمد برای تشخیص بهترین پاسخ‌ها در سیستم‌های پرسش و پاسخ آنلاین با استفاده از روش‌های یادگیری جمعی و تکنیک‌های نوین پردازش زبان طبیعی پیشنهاد شد. پس از پیاده‌سازی و ارزیابی مدل‌های مختلف، نتایج نشان داد که مدل پیشنهادی توانسته است دقت قابل توجهی در تشخیص پاسخ‌های صحیح داشته باشد، اما همچنان چالش‌هایی برای بهبود نتایج وجود دارد. از جمله مهم‌ترین مشکلات، وجود خطاهای FP و FN بود که به دلیل شباهت در درصد وقوع پاسخ‌ها یا انتخاب پاسخ‌هایی با درصد وقوع مشابه به عنوان بهترین پاسخ به وجود می‌آمد.

در فرآیند این تحقیق، مدل‌های مختلفی از جمله XLNet و RoBERTa برای استخراج ویژگی‌های متنی و جنگل‌های تصادفی برای استخراج ویژگی‌های غیرمتنی پیاده‌سازی شدند. در نهایت، با استفاده از روش رأی‌گیری اکثریت و مدل فرا یادگیرنده درخت تصمیم، دقت نهایی به ۹۹.۰۳ درصد رسید. تغییراتی مانند استفاده از خروجی نرم مدل فرا یادگیرنده به جای پیش‌بینی قطعی منجر به بهبود عملکرد در برخی از سوالات شد، به طوری که ۲۹ سوال که در مدل قبلی خطا داشتند، بهبود یافتند.

با این حال، با وجود افزایش دقت، چالش‌هایی مانند انتخاب چندین پاسخ به عنوان بهترین پاسخ یا انتخاب پاسخی که به طور ظاهری بهترین بود اما از نظر کاربر مناسب نبود، همچنان باقی ماند. برای بهبود بیشتر، پیشنهاد می‌شود که خروجی یادگیرنده‌های پایه نیز به صورت درصد وقوع تنظیم شود و این درصدها به مدل فرا یادگیرنده درخت تصمیم داده شود. این اصلاحات می‌تواند به کاهش خطاها کمک کند و دقت مدل را افزایش دهد.

در نهایت، این پژوهش نشان داد که استفاده از مدل‌های پیچیده و یادگیری جمعی در سیستم‌های پرسش و پاسخ می‌تواند به عملکرد بهتری در تشخیص پاسخ‌های صحیح منجر شود، اما نیاز به اصلاحات بیشتر برای رفع چالش‌ها و بهبود دقت دارد. این دستاوردها، زمینه‌ساز تحقیقات آینده در جهت بهبود مدل‌ها و رفع مشکلات موجود خواهند بود.

۷- پیشنهادها برای بهبود مدل و تحقیقات آتی

در این پژوهش، مدل یادگیری جمعی پشته‌ای وزن‌دار با استفاده از مدل‌های پایه XLNet، RoBERTa، جنگل‌های تصادفی و رأی‌گیری اکثریت، همراه با فرایادگیر درخت تصمیم مورد بررسی قرار گرفت. اگرچه این روش توانست در بسیاری از موارد عملکرد مناسبی داشته باشد، اما همچنان چالش‌ها و محدودیت‌هایی در شناسایی بهترین پاسخ‌ها وجود داشت. بر این اساس، پیشنهادهایی برای بهبود عملکرد مدل و تحقیقات آتی ارائه می‌شود.

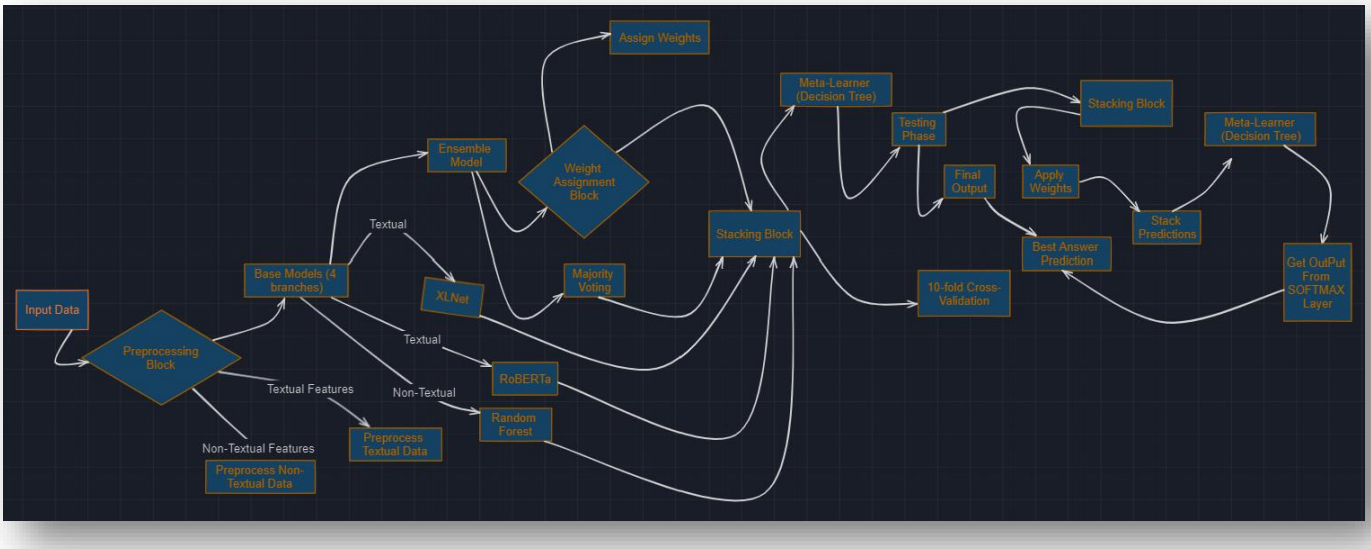
یکی از مهم‌ترین چالش‌های شناسایی‌شده، استفاده از پیش‌بینی قطعی کلاس در فرایادگیر درخت تصمیم بود. نتایج نشان داد که استفاده از درصد وقوع کلاس‌ها به‌جای پیش‌بینی قطعی می‌تواند عملکرد مدل را بهبود بخشد و از بروز خطاهای ناشی از همپوشانی احتمالی جلوگیری کند. همچنین، در روش رأی‌گیری اکثریت، وزن‌دهی ثابت مورد استفاده قرار گرفت که ممکن است در شرایط مختلف عملکرد بهینه‌ای نداشته باشد. پیشنهاد می‌شود از رویکرد وزن‌دهی پویا استفاده شود، به‌طوری که وزن هر مدل پایه بر اساس عملکرد آن در لحظه تنظیم شود. این تغییر می‌تواند به بهبود قابل‌توجه دقت مدل منجر شود.

تحلیل دقیق‌تر خطاها و استفاده از تکنیک‌های تقویت یادگیری نظیر AdaBoost و Gradient Boosting نیز می‌تواند به کاهش خطاها و یادگیری از اشتباهات گذشته کمک کند. علاوه بر این، گسترش مجموعه داده‌ها با نمونه‌های متنوع‌تر و افزایش تعداد داده‌های آموزشی، به ویژه با افزودن سوالات نامشخص و بدون پاسخ بهینه، می‌تواند مدل را در مواجهه با شرایط واقعی قدرتمندتر سازد. اضافه کردن چنین داده‌هایی، چالش‌های بیشتری برای مدل ایجاد کرده و توانایی آن را در مدیریت شرایط نامطمئن افزایش می‌دهد.

یکی دیگر از رویکردهای پیشنهادی، استفاده از یادگیری انتقالی است. این روش شامل آموزش اولیه مدل بر روی مجموعه داده‌های گسترده‌تر و حاوی اطلاعات پیش‌زمینه‌ای مرتبط است که می‌تواند دانش عمیق‌تری به مدل اضافه کند و توانایی آن را در شناسایی دقیق بهترین پاسخ‌ها افزایش

دهد. برای اطمینان از دقت مدل و کیفیت پاسخ‌های ارائه‌شده، پیشنهاد می‌شود پاسخ‌ها به‌صورت دستی و بر اساس ۱۳ معیار اصلی ارزیابی شوند که شامل آموزنده بودن، ادب، جامعیت، خوانایی، ارتباط مستقیم با سوال، اختصار، قانع‌کنندگی، جزئیات دقیق، اصالت، واقع‌گرایی، نوآوری، مفید بودن، و کارشناسی‌شدن است. تحلیل نتایج این ارزیابی‌ها با استفاده از ضریب کاپا کوهن می‌تواند میزان تطابق عملکرد مدل با ارزیابی انسانی را به‌طور دقیق مشخص کند و تصویری شفاف از کیفیت پاسخ‌های تولیدشده ارائه دهد.

در نهایت، برای کاهش تعصبات در سیستم‌های پرسش و پاسخ، الگوریتم‌های کالیبراسیون رأی می‌توانند به‌کار گرفته شوند. این الگوریتم‌ها با در نظر گرفتن ویژگی‌های رأی‌دهندگان و تعاملات آن‌ها، وزن‌دهی رأی‌ها را بهبود بخشیده و دقت فرآیند انتخاب بهترین پاسخ‌ها را افزایش می‌دهند. اجرای این پیشنهادها می‌تواند به رفع محدودیت‌های کنونی و افزایش کارایی و دقت مدل در شناسایی بهترین پاسخ‌ها منجر شود.



شکل ۷: بلوک دیاگرام کلی مدل

مراجع

[۱] Mamykina, L.; Manoim, B.; Mittal, M.; Hripcsak, G.; Hartmann, B. Design lessons from the fastest q&a site in the west. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ۲۰۱۱; pp. ۲۸۵۷–۲۸٦٦.

[۲] Gkotsis, G.; Liakata, M.; Pedrinaci, C.; Stepanyan, K.; Domingue, J. ACQUA: automated community-based question answering through the discretisation of shallow linguistic features. J. Web Sci. ۲۰۱۵, ۱, ۱–۱۵.

[۳] Rajpurkar, P.; Zhang, J.; Lopyrev, K.; Liang, P. Squad: ۱۰۰,۰۰۰+ questions for machine comprehension of text. arXiv ۲۰۱٦, arXiv:۱٦۰٦,۰٥۲٥۰.

[٤] Cui, W.; Xiao, Y.; Wang, H.; Song, Y.; Hwang, S.w.; Wang, W. KBQA: learning question answering over QA corpora and knowledge bases. arXiv ۲۰۱۹, arXiv:۱۹۰۳,۰۲٤۱۹.

[٥] Zhou, M.; Shi, Z.; Huang, M.; Zhu, X. Knowledge-Aided Open-Domain Question Answering. arXiv ۲۰۲۰, arXiv:۲۰۰٦,۰٥۲٤٤.

[٦] Correa, D., & Sureka, A. (۲۰۱٤, April). Chaff from the wheat: Characterization and modeling of deleted questions on StackOverflow. In Proceedings of the ۲۳rd international conference on World wide web (pp. ٦۳۱-٦٤۲).

[۷] Jeon, J., Croft, W. B., Lee, J. H., & Park, S. (۲۰۰٦, August). A framework to predict the quality of answers with non-textual features. In Proceedings of the ۲۹th annual international ACM SIGIR conference on Research and development in information retrieval (pp. ۲۲۸-۲۳٥).

[۸] Shah, C., & Pomerantz, J. (۲۰۱۰, July). Evaluating and predicting answer quality in community QA. In Proceedings of the ۳۳rd international ACM SIGIR conference on Research and development in information retrieval (pp. ٤۱۱-٤۱۸).

[۹] Ponzanelli, L., Mocci, A., Bacchelli, A., Lanza, M., & Fullerton, D. (۲۰۱٤, September). Improving low quality StackOverflow post detection. In ۲۰۱٤ IEEE international conference on software maintenance and evolution (pp. ٥٤۱-٥٤٤). IEEE.

[۱۰] Chen, A., Stanovsky, G., Singh, S., & Gardner, M. (۲۰۱۹, November). Evaluating question answering evaluation. In Proceedings of the ۲nd workshop on machine reading for question answering (pp. ۱۱۹-۱۲٤).

[۱۱] Hu, Z., Zhang, Z., Yang, H., Chen, Q., & Zuo, D. (۲۰۱۷). A deep learning approach for predicting the quality of online health expert question-answering services. Journal of biomedical informatics, ۷۱, ۲٤۱-۲٥۳.

[۱۲] Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (۲۰۰۸, February). Finding high-quality content in social media. In Proceedings of the ۲۰۰۸ international conference on web search and data mining (pp. ۱۸۳-۱۹٤).

[۱۳] Blooma, M. J., Chua, A. Y., & Goh, D. H. L. (۲۰۰۸, March). A predictive framework for retrieving the best answer. In Proceedings of the ۲۰۰۸ ACM symposium on Applied computing (pp. ۱۱۰۷-۱۱۱۱).

[۱٤] Suggu, S. P., Goutham, K. N., Chinnakotla, M., & Shrivastava, M. (۲۰۱٦, December). Hand in glove: Deep feature fusion network architectures for answer quality prediction in community question answering. In Proceedings of COLING ۲۰۱٦, the ۲٦th International Conference on Computational Linguistics: Technical Papers (pp. ۱٤۲۹-۱٤٤۰).

[۱٥] Hadfi, R., Moustafa, A., Yoshino, K., & Ito, T. (۲۰۲۲). Best-Answer Prediction in Q&A Sites Using User Information. arXiv preprint arXiv:۲۲۱۲,۰۸٤۷٥.

[۱٦] Gkotsis, G., Liakata, M., Pedrinaci, C., & Domingue, J. (۲۰۱٥). Leveraging Textual Features for Best Answer Prediction in Community-based Question Answering. arXiv preprint arXiv:۱٥۰٦,۰۲۸۱٦.

[۱۷] Zhu, Z., Bernhard, D., & Gurevych, I. (۲۰۰۹, November). A multi-dimensional model for assessing the quality of answers in social Q&A sites. In ICIQ (pp. ۲٦٤-۲٦٥).

[۱۸] Liu, M., Liu, Y., & Yang, Q. (۲۰۱۰). Predicting best answerers for new questions in community question answering. In Web-Age Information Management: ۱۱th International Conference, WAIM ۲۰۱۰, Jiuzhaigou, China, July ۱٥-۱۷, ۲۰۱۰. Proceedings ۱۱ (pp. ۱۲۷-۱۳۸). Springer Berlin Heidelberg.

[۱۹] Blooma, M. J., Chua, A. Y. K., & Goh, D. H. L. (۲۰۱۰, April). Selection of the best answer in CQA services. In ۲۰۱۰ Seventh International Conference on Information Technology: New Generations (pp. ٥۳٤-٥۳۹). IEEE.

[۲۰] Burel, G., He, Y., & Alani, H. (۲۰۱۲). Automatic identification of best answers in online enquiry communities. In The Semantic Web: Research and Applications: ۹th Extended Semantic Web Conference, ESWC ۲۰۱۲, Heraklion, Crete, Greece, May ۲۷-۳۱, ۲۰۱۲. Proceedings ۹ (pp. ٥۱٤-٥۲۹). Springer Berlin Heidelberg.

[۲۱] Gkotsis, G., Stepanyan, K., Pedrinaci, C., Domingue, J., & Liakata, M. (۲۰۱٤, June). It's all in the content: state of the art best answer prediction based on discretisation of shallow linguistic features. In Proceedings of the ۲۰۱٤ ACM conference on Web science (pp. ۲۰۲-۲۱۰).

[۲۲] Chen, B. C., Dasgupta, A., Wang, X., & Yang, J. (۲۰۱۲, August). Vote calibration in community question-answering systems. In Proceedings of the ۳٥th international ACM SIGIR conference on Research and development in information retrieval (pp. ۷۸۱-۷۹۰).

[۲۳] Elalfy, D., Gad, W., & Ismail, R. (۲۰۱۸). A hybrid model to predict best answers in question answering communities. Egyptian informatics journal, ۱۹(۱), ۲۱-۳۱.

[۲٤] Zheng, W., & Li, M. (۲۰۱۷). The best answer prediction by exploiting heterogeneous data on software development Q&A forum. Neurocomputing, ۲٦۹, ۲۱۲-۲۱۹.

[۲٥] Tondulkar, R., Dubey, M., & Desarkar, M. S. (۲۰۱۸, September). Get me the best: predicting best answerers in community question answering sites. In Proceedings of the ۱۲th ACM Conference on Recommender Systems (pp. ۲٥۱-۲٥۹).

[۲٦] Chen, D., Fisch, A., Weston, J., & Bordes, A. (۲۰۱۷). Reading wikipedia to answer open-domain questions. arXiv preprint arXiv:۱۷۰٤,۰۰۰٥۱.

[۲۷] Hu, X. (۲۰۲۳). Enhancing Answer Selection in Community Question Answering with Pre-trained and Large Language Models. arXiv preprint arXiv:۲۳۱۱,۱۷٥۰۲.

[۲۸] Toba, H., Ming, Z. Y., Adriani, M., & Chua, T. S. (۲۰۱٤). Discovering high quality answers in community question answering archives using a hierarchy of classifiers. Information Sciences, ۲٦۱, ۱۰۱-۱۱٥.

[۲۹] Calefato, F., Lanubile, F., & Novielli, N. (۲۰۱۹). An empirical assessment of best-answer prediction models in technical Q&A sites. Empirical Software Engineering, ۲٤, ۸٥٤-۹۰۱.

[۳۰] Gao, Z., Xia, X., Lo, D., & Grundy, J. (۲۰۲۰). Technical Q۸A site answer recommendation via question boosting. ACM Transactions on Software Engineering and Methodology (TOSEM), ۳۰(۱), ۱-۳٤.

[۳۱] Lou, R., Zhang, K., & Yin, W. (۲۰۲۳). A comprehensive survey on instruction following. arXiv preprint arXiv:۲۳۰۳,۱۰٤۷٥.

[۳۲] Sakai, T., Ishikawa, D., Kando, N., Seki, Y., Kuriyama, K., & Lin, C. Y. (۲۰۱۱, February). Using graded-relevance metrics for evaluating community QA answer selection. In Proceedings of the fourth ACM international conference on Web search and data mining (pp. ۱۸۷-۱۹٦).

[۳۳] Valentin, L. (۲۰۲۲). Answer ranking in Community Question Answering: a deep learning approach. arXiv preprint arXiv:۲۲۱۲,۰۱۲۱۸.

[۳٤] Roy, P. K., Saumya, S., Singh, J. P., Banerjee, S., & Gutub, A. (۲۰۲۳). Analysis of community question-answering issues via machine learning and deep learning: State-of-the-art review. CAAI Transactions on Intelligence Technology, ۸(۱), ۹٥-۱۱۷.

[۳٥] Roy, P. K., Ahmad, Z., Singh, J. P., Alryalat, M. A. A., Rana, N. P., & Dwivedi, Y. K. (۲۰۱۸). Finding and ranking high-quality answers in community question answering sites. Global Journal of Flexible Systems Management, ۱۹, ٥۳-٦۸.

[۳٦] Zoratto, V., Godoy, D., & Aranda, G. N. (۲۰۲۳). A Study on Influential Features for Predicting Best Answers in Community Question-Answering Forums. Information, ۱٤(۹), ٤۹٦.

[۳۷] Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (۲۰۱۱, July). Optimizing semantic coherence in topic models. In Proceedings of the ۲۰۱۱ conference on empirical methods in natural language processing (pp. ۲٦۲-۲۷۲).

[۳۸] Molino, P., Aiello, L. M., & Lops, P. (۲۰۱٦). Social question answering: Textual, user, and network features for best answer prediction. ACM Transactions on Information Systems (TOIS), ۳٥(۱), ۱-٤۰.

[۳۹] Anbaraki, S., & Jowkar, A. (۲۰۲۱). Evaluating and Predicting the Quality of Answers Factors in the Research Gateâs Question and Answer System: a Case Study of the Thematic Domain of Knowledge Management. Iranian Journal of Information Processing and Management, ۳٦(۳), ۷۰۹-۷۳٦. doi: ۱۰,٥۲٥٤۷/jipm.۳٦,۳,۷۰۹